

# Estimating the Division Kernel of a Size-Structured Population

Van Ha Hoang \*

May 20, 2016

## Abstract

We consider a size-structured population describing the cell divisions. The cell population is described by an empirical measure and we observe the divisions in the continuous time interval  $[0, T]$ . We address here the problem of estimating the division kernel  $h$  (or fragmentation kernel) in case of complete data. An adaptive estimator of  $h$  is constructed based on a kernel function  $K$  with a fully data-driven bandwidth selection method. We obtain an oracle inequality and an exponential convergence rate, for which optimality is considered.

**Keywords:** random size-structured population, division kernel, nonparametric estimation, Goldenshluger-Lepski's method, adaptive estimator, penalization, optimal rate.

## 1 Introduction

Models for populations of dividing cells possibly differentiated by covariates such as size have made the subject of an abundant literature in recent years (starting from Athreya and Ney [3], Harris [18], Jagers [28]...) Covariates termed as ‘size’ are variables that grow deterministically with time (such as volume, length, level of certain proteins, DNA content, *etc.*) Such models of structured populations provide descriptions for the evolution of the size distribution, which can be interesting for applications. For instance, in the spirit of Stewart *et al.* [33], we can imagine that each cell contains some toxicities whose quantity plays the role of the size. The asymmetric divisions of the cells, where one daughter contains more toxicity than the other, can lead under some conditions to the purge of the toxicity in the population by concentrating it into few lineages. These results are linked with the concept of aging for cell lineage. This concept has been tackled in many papers (e.g. Ackermann *et al.* [2], Aguilaniu *et al.* [1], C-Y. Lai *et al.* [22], Evans and Steinsaltz [15], Moseley [27]...).

Here we consider a stochastic individual-based model of size-structured population in continuous time, where individuals are cells undergoing asymmetric binary divisions and whose size is the quantity of toxicity they contain. A cell containing a toxicity  $x \in \mathbb{R}_+$  divides at a rate  $R > 0$ . The toxicity grows inside the cell with rate  $\alpha > 0$ . When a cell divides, a random fraction  $\Gamma \in [0, 1]$  of the toxicity goes in the first daughter cell and

---

\*Laboratoire Paul Painlevé UMR CNRS 8524, Université Lille 1.

Email address: van-ha.hoang@math.univ-lille1.fr

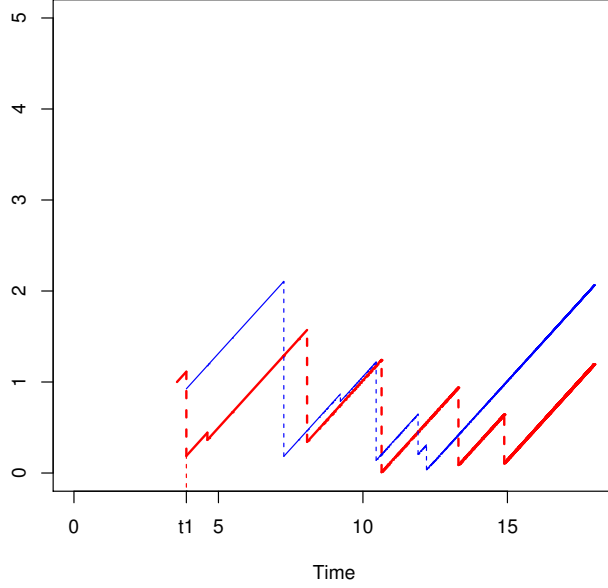


Figure 1: Trajectories of two daughter cells after a division, separating after the first division at time  $t_1$ .

$1 - \Gamma$  in the second one. If  $\Gamma = \frac{1}{2}$ , the daughters are the same with toxicity  $\frac{x}{2}$ . We assume that  $\Gamma$  has a symmetric distribution on  $[0, 1]$  with a density  $h$  with respect to Lebesgue measure such that  $\mathbb{P}(\Gamma = 0) = \mathbb{P}(\Gamma = 1) = 0$ . If  $h$  is peaked at  $1/2$  (*i.e.*  $\Gamma \simeq 1/2$ ), then both daughters contain the same toxicity, *i.e.* the half of their mother's toxicity. The more  $h$  puts weight in the neighbourhood of 0 and 1, the more asymmetric the divisions are, with one daughter having little toxicity and the other a toxicity close to its mother's one. If we consider that having a lot of toxicity is a kind of senescence, then, the kurtosis of  $h$  provides indication on aging phenomena (see [24]).

Modifications of this model to account for more complex phenomena have been considered in other papers. Bansaye and Tran [6], Cloez [11] or Tran [35] consider non-constant division and growth rates. Robert *et al.* [29] studies whether divisions can occur only when a size threshold is reached. Our purpose here is to estimate the density  $h$  ruling the divisions, and we stick to constant rates  $R$  and  $\alpha$  for the sake of simplicity. Notice that several similar models for binary cell division in discrete time also exist in the literature and have motivated statistical question as here, see for instance Bansaye *et al.* [4, 7], Bercu *et al.* [8], Bitseki Penda [10], Delmas and Marsalle [12] or Guyon [17].

Individual-based models provide a natural framework for statistical estimation. Estimation of the division rate is, for instance, the subject of Doumic *et al.* [13, 14] and Hoffmann and Olivier [19]. Here, the density  $h$  is the kernel division that we want to estimate. Assuming that we observe the divisions of cells in continuous time on the interval  $[0, T]$ , with  $T > 0$ , we propose an adaptive kernel estimator  $\hat{h}$  of  $h$  for which we obtain an oracle inequality in Theorem 2. The construction of  $\hat{h}$  is detailed in the sequel. From oracle inequality we can infer adaptive exponential rates of convergence with respect to  $T$  depending on  $\beta$  the smoothness of the density. Most of the time, nonparametric rates are of the form  $n^{-\frac{2\beta}{2\beta+1}}$  (see for instance Tsybakov [36]) and exponential rates are not

often encountered in the literature. The exponential rates are due to binary splitting, the number of cells *i.e* the sample size increases exponentially in  $\exp(RT)$  (see Section 2.3). By comparison, in [19] Hoffmann and Olivier obtain a similar rate of convergence  $\exp\left(-\lambda_B \frac{\varsigma}{2\varsigma+1} T\right)$  of the kernel estimator of their division rate  $B(x)$ , where  $\lambda_B$  is the Malthus parameter and  $\varsigma > 0$  is the smoothness of  $B(x)$ . However, their estimator  $\hat{B}_T$  of  $B$  is not adaptive since the choice of their optimal bandwidth still depends on  $\varsigma$ . Our estimator is adaptive with an “optimal” bandwidth chosen from a data-driven method. We derive upper bounds and lower bounds for asymptotic minimax risks on Hölder classes and show that they coincide. Hence, the rate of convergence of our estimator  $\hat{h}$  proves to be optimal in the minimax sense on the Hölder classes.

This paper is organized as follows. In Section 2, we introduce a stochastic differential equation driven by a Poisson point measure to describe the population of cells. Then, we construct the estimator of  $h$  and obtain upper and lower bounds for the MISE (Mean Integrated Squared Error). Our main results are stated in Theorems 3 and 4. Numerical results and discussions about aging effect are presented in Section 3. The main proofs are shown in Section 4.

**Notation** We introduce some notations used in the sequel.

Hereafter,  $\|\cdot\|_1$  and  $\|\cdot\|_2$  denote the  $\mathbb{L}^1$  and  $\mathbb{L}^2$  norms on  $\mathbb{R}$  with respect to Lebesgue measure:

$$\|f\|_1 = \int_{\mathbb{R}} |f(\gamma)| d\gamma, \quad \|f\|_2 = \left( \int_{\mathbb{R}} |f(\gamma)|^2 d\gamma \right)^{1/2}.$$

The  $\mathbb{L}^\infty$  norm is defined by

$$\|f\|_\infty = \sup_{\gamma \in (0,1)} |f(\gamma)|.$$

Finally,  $f \star g$  denotes the convolution of two functions  $f$  and  $g$  defined by

$$f \star g(\gamma) = \int_{\mathbb{R}} f(u) g(\gamma - u) du.$$

## 2 Microscopic model and kernel estimator of $h$

### 2.1 The model

We recall the Ulam-Harris-Neveu notation used to describe the genealogical tree. The first cell is labelled by  $\emptyset$  and when the cell  $i$  divides, the two descendants are labelled by  $i0$  and  $i1$ . The set of labels is

$$J = \{\emptyset\} \cup \bigcup_{m=1}^{\infty} \{0,1\}^m. \quad (1)$$

We denote  $V_t$  the set of cells alive at time  $t$ , and  $V_t \subset J$ .

Let  $\mathcal{M}_F(\mathbb{R}_+)$  be the space of finite measures on  $\mathbb{R}_+$  embedded with the topology of weak convergence and  $X_t^i$  be the quantity of toxicity in the cell  $i$  at time  $t$ , we describe the population of cells at time  $t$  by a random point measure in  $\mathcal{M}_F(\mathbb{R}_+)$ :

$$Z_t(dx) = \sum_{i=1}^{N_t} \delta_{X_t^i}(dx), \quad \text{where} \quad N_t = \langle Z_t, 1 \rangle = \int_{\mathbb{R}_+} Z_t(dx) \quad (2)$$

is the number of individuals living at time  $t$ . For a measure  $\mu \in \mathcal{M}_F(\mathbb{R}_+)$  and a positive function  $f$ , we use the notation  $\langle \mu, f \rangle = \int_{\mathbb{R}_+} f d\mu$ .

Along branches of the genealogical tree, the toxicity  $(X_t, t \geq 0)$  satisfies

$$dX_t = \alpha dt, \quad (3)$$

with  $X_0 = x_0$ . When the cells divide, the toxicity is shared between the daughter cells. This is described by the following stochastic differential equation (SDE).

Let  $Z_0 \in \mathcal{M}_F(\mathbb{R}_+)$  be an initial condition such that

$$\mathbb{E}(\langle Z_0, 1 \rangle) < +\infty, \quad (4)$$

and let  $Q(ds, di, d\gamma)$  be a Poisson point measure on  $\mathbb{R}_+ \times \mathcal{E} := \mathbb{R}_+ \times J \times [0, 1]$  with intensity  $q(ds, di, d\gamma) = R ds n(di) H(d\gamma)$ .  $n(di)$  is the counting measure on  $J$  and  $ds$  is Lebesgue measure on  $\mathbb{R}_+$ . We denote  $\{\mathcal{F}_t\}_{t \geq 0}$  the canonical filtration associated with the Poisson point measure and the initial condition. The stochastic process  $(Z_t)_{t \geq 0}$  can be described by a SDE as follows.

**Definition 1.** *For every test function  $f_t(x) = f(x, t) \in \mathcal{C}_b^{1,1}(\mathbb{R}_+ \times \mathbb{R}_+, \mathbb{R})$  (bounded of class  $\mathcal{C}^1$  in  $t$  and  $x$  with bounded derivatives), the population of cells is described by:*

$$\begin{aligned} \langle Z_t, f_t \rangle &= \langle Z_0, f_0 \rangle + \int_0^t \int_{\mathbb{R}_+} (\partial_s f_s(x) + \alpha \partial_x f_s(x)) Z_s(dx) ds \\ &+ \int_0^t \int_{\mathcal{E}} \mathbb{1}_{\{i \leq N_{s-}\}} \left[ f_s \left( \gamma X_{s-}^i \right) + f_s \left( (1 - \gamma) X_{s-}^i \right) - f_s \left( X_{s-}^i \right) \right] Q(ds, di, d\gamma). \end{aligned} \quad (5)$$

The second term in the right hand side of (5) corresponds to the growth of toxicities in the cells and the third term gives a description of cell divisions where the sharing of toxicity into two daughter cells depends on the random fraction  $\Gamma$ .

We now state some properties of  $N_t$  that are useful in the sequel.

**Proposition 1.** *Let  $T > 0$ , and assume the initial condition  $N_0$ , the number of mother cells at time  $t = 0$ , is deterministic, for the sake of simplicity. We have*

i) *Let  $T_j$  be the  $j^{\text{th}}$  jump time. Then:*

$$\lim_{j \rightarrow +\infty} T_j = +\infty \text{ and } \lim_{T \rightarrow +\infty} N_T = +\infty \quad (a.s). \quad (6)$$

ii)  *$N_T$  is distributed according to a negative binomial distribution, denoted as  $\mathcal{NB}(N_0, e^{-RT})$ . Its probability mass function is then*

$$\mathbb{P}(N_T = n) = \binom{n-1}{n-N_0} (e^{-RT})^{N_0} (1 - e^{-RT})^{n-N_0}, \quad (7)$$

for  $n \geq N_0$ . When  $N_0 = 1$ ,  $N_T$  has a geometric distribution

$$\mathbb{P}(N_T = n) = e^{-RT} (1 - e^{-RT})^{n-1}. \quad (8)$$

Consequently, we have

$$\mathbb{E}[N_T] = N_0 e^{RT}. \quad (9)$$

iii) When  $N_0 = 1$ :

$$\mathbb{E} \left[ \frac{1}{N_T} \right] = \frac{RTe^{-RT}}{1 - e^{-RT}}. \quad (10)$$

When  $N_0 > 1$ , we have:

$$\mathbb{E} \left[ \frac{1}{N_T} \right] = \left( \frac{e^{-RT}}{1 - e^{-RT}} \right)^{N_0} (-1)^{N_0-1} \left( \sum_{k=1}^{N_0-1} \binom{N_0-1}{k} \frac{(-1)^k e^{kRT}}{k} + RT \right). \quad (11)$$

iv) Furthermore, when  $N_0 > 1$ , we have

$$\frac{e^{-RT}}{N_0} \leq \mathbb{E} \left[ \frac{1}{N_T} \right] \leq \frac{e^{-RT}}{N_0 - 1}. \quad (12)$$

The proof of Proposition 1 is presented in Section 4.

## 2.2 Influence of age

In this section, we study the aging effect via the mean age which is defined as follows.

**Definition 2.** The mean age of the cell population up to time  $t \in \mathbb{R}_+$  is defined by:

$$\bar{X}_t = \frac{1}{N_t} \sum_{i=1}^{N_t} X_t^i = \frac{\langle Z_t, f \rangle}{N_t}, \quad (13)$$

where  $f(x) = x$ .

Following the work of Bansaye et al. [5], we note that the long time behavior of the mean age is related to the law of an auxiliary process  $Y$  started at  $Y_0 = \frac{X_0}{N_0}$  with infinitesimal generator characterized for all  $f \in \mathcal{C}_b^{1,1}(\mathbb{R}_+, \mathbb{R})$  by

$$Af(x) = \alpha f'(x) + 2R \int_0^1 (f(\gamma x) - f(x)) h(\gamma) d\gamma. \quad (14)$$

The empirical distribution  $\frac{1}{N_t} \sum_{i=1}^{N_t} \delta_{X_t^i}$  gives the law of the path of a particle chosen at random at time  $t$ . Heuristically, the distribution of  $Y$  restricted to  $[0, t]$  approximates this distribution. Hence, this explains the coefficient 2 which is a size-biased phenomenon, *i.e.* when one chooses a cell in the population at time  $t$ , a cell belonging to a branch with more descendants is more likely to be chosen.

**Lemma 1.** Let  $Y$  be the auxiliary process with infinitesimal generator (14), for  $t \in \mathbb{R}_+$ ,

$$Y_t = \left( Y_0 - \frac{\alpha}{R} \right) e^{-Rt} + \frac{\alpha}{R} + \int_0^t e^{-R(t-s)} dU_s. \quad (15)$$

where  $U_t$  is a square-integrable martingale.

Consequently, we have

$$\mathbb{E}[Y_t] = \left( Y_0 - \frac{\alpha}{R} \right) e^{-Rt} + \frac{\alpha}{R}, \quad (16)$$

and

$$\lim_{t \rightarrow \infty} \mathbb{E}[Y_t] = \frac{\alpha}{R}. \quad (17)$$

We will show that the auxiliary process  $Y$  satisfies ergodic properties (see Section 4.3) which entails the following theorem.

**Theorem 1.** *Assume that there exists  $\underline{h} > 0$  such that for all  $\gamma \in (0, 1)$ ,  $h(\gamma) \geq \underline{h}$ . Then*

$$\lim_{t \rightarrow +\infty} \bar{X}_t = \lim_{t \rightarrow +\infty} \mathbb{E}(Y_t) = \frac{\alpha}{R}. \quad (18)$$

Theorem 1 is a consequence of the ergodic properties of  $Y$ , of Theorem 4.2 in Bansaye et al. [5] and of Lemma 1. It shows that the average of the mean age tends to the constant  $\alpha/R$  when the time  $t$  is large. Simulations in Section 3 illustrate the results. The proofs of Lemma 1 and Theorem 1 are presented in Section 4.2 and Section 4.3.

**Remark 1.** *When the population is large, we are interested in studying the asymptotic behavior of the random point measure. As in Doumic et al. [14], we can show that our stochastic model is approximated by a growth-fragmentation partial differential equation. This problem is a work in progress.*

## 2.3 Estimation of the division kernel

### Data and construction of the estimator

Suppose that we observe the evolution of the cell population in a given time interval  $[0, T]$ . At the  $i^{\text{th}}$  division time  $t_i$ , let us denote  $j_i$  the individual who splits into two daughters  $X_{t_i}^{j_i 0}$  and  $X_{t_i}^{j_i 1}$  and define

$$\Gamma_i^0 = \frac{X_{t_i}^{j_i 0}}{X_{t_i}^{j_i}} \quad \text{and} \quad \Gamma_i^1 = \frac{X_{t_i}^{j_i 1}}{X_{t_i}^{j_i}},$$

the random fractions that go into the daughter cells, with the convention  $\frac{0}{0} = 0$ .

$\Gamma_i^0$  and  $\Gamma_i^1$  are exchangeable with  $\Gamma_i^0 + \Gamma_i^1 = 1$ ,  $\Gamma_i^0$  and  $\Gamma_i^1$  are thus not independent but the couples  $(\Gamma_i^0, \Gamma_i^1)_{i \in \mathbb{N}^*}$  are independent and identically distributed with distribution  $(\Gamma^0, \Gamma^1)$  where  $\Gamma^1 \sim H(d\gamma)$  and  $\Gamma^0 = 1 - \Gamma^1$ .

Since  $h$  is a density function, it is natural to use a kernel method. We define an estimator  $\hat{h}_\ell$  of  $h$  based on the data  $(\Gamma_i^0, \Gamma_i^1)_{i \in \mathbb{N}^*}$  as follows.

**Definition 3.** *Let  $K : \mathbb{R} \rightarrow \mathbb{R}$  is an integrable function such that*

$$\int_{\mathbb{R}} K(x) dx = 1 \quad \text{and} \quad \int_{\mathbb{R}} K^2(x) dx < \infty.$$

*Let  $M_T$  be the random number of divisions in the time interval  $[0, T]$  and assume that  $M_T > 0$ . For all  $\gamma \in (0, 1)$ , define*

$$\hat{h}_\ell(\gamma) = \frac{1}{M_T} \sum_{i=1}^{M_T} K_\ell(\gamma - \Gamma_i^1), \quad (19)$$

*where  $K_\ell = \frac{1}{\ell} K(\cdot/\ell)$ ,  $\ell > 0$  is the bandwidth to be chosen.*

**Remark 2.** *Since  $N_0 \neq 0$ , the number of random divisions  $M_T$  is not equal to the number of individuals living at time  $T$ . Indeed, we have  $M_T = N_T - N_0$ .*

In (19),  $\hat{h}_\ell$  depends also on  $T$ . However, we omit  $T$  for the sake of notation. The estimator  $\hat{h}_\ell$  will satisfy the following properties.

**Proposition 2.**

*i) The conditional expectation and conditional variance given  $M_T$  of  $\hat{h}_\ell(\gamma)$  and variance  $\hat{h}_\ell(\gamma)$  are:*

$$\mathbb{E}[\hat{h}_\ell(\gamma)|M_T] = K_\ell \star h(\gamma) \text{ and } \mathbb{E}[\hat{h}_\ell(\gamma)] = K_\ell \star h(\gamma), \quad (20)$$

$$\mathbb{V}ar[\hat{h}_\ell(\gamma)|M_T] = \frac{1}{M_T} \mathbb{V}ar \left[ K_\ell(\gamma - \Gamma_1^1) \right], \quad (21)$$

$$\mathbb{V}ar[\hat{h}_\ell(\gamma)] = \mathbb{E} \left[ \frac{1}{M_T} \right] \mathbb{V}ar \left[ K_\ell(\gamma - \Gamma_1^1) \right]. \quad (22)$$

Consequently, we have  $\mathbb{E}[\hat{h}_\ell(\gamma)|M_T] = \mathbb{E}[\hat{h}_\ell(\gamma)]$ .

*ii) For all  $\gamma \in (0, 1)$ ,*

$$\lim_{T \rightarrow +\infty} \hat{h}_\ell(\gamma) = K_\ell \star h(\gamma) \quad (a.s). \quad (23)$$

**Adaptive estimation of  $h$  by Goldenshluger and Lepski's (GL) method**

Let  $\hat{h}_\ell$  be the kernel estimator of  $h$  as in Definition 3. We measure the performance of  $\hat{h}_\ell$  via its  $\mathbb{L}^2$ -loss *i.e* the average  $\mathbb{L}^2$  distance between  $\hat{h}_\ell$  and  $h$ . The objective is to find a bandwidth which minimizes this  $\mathbb{L}^2$ -loss. Since  $M_T$  is random, we first study the  $\mathbb{L}^2$ -loss conditionally to  $M_T$ .

**Proposition 3.** *The  $\mathbb{L}^2$ -loss of  $\hat{h}_\ell$  given  $M_T$  satisfies :*

$$\mathbb{E} \left[ \|\hat{h}_\ell - h\|_2 | M_T \right] \leq \|h - K_\ell \star h\|_2 + \frac{\|K\|_2}{\sqrt{M_T \ell}}. \quad (24)$$

In the right hand side of the risk decomposition (24) the first term is a bias term. Hence it decreases when  $\ell \rightarrow 0$  whereas the second term which is a variance term increases when  $\ell \rightarrow 0$ . The best choice of  $\ell$  should minimize this bias-variance trade-off. Thus, from a finite family of bandwidths  $H$ , the best bandwidth  $\bar{\ell}$  would be

$$\bar{\ell} := \operatorname{argmin}_{\ell \in H} \left\{ \|h - K_\ell \star h\|_2 + \frac{\|K\|_2}{\sqrt{M_T \ell}} \right\}. \quad (25)$$

The bandwidth  $\bar{\ell}$  is called "the oracle bandwidth" since it depends on  $h$  which is unknown and then it cannot be used in practice. Since the oracle bandwidth minimizes a bias variance trade-off, we need to find an estimation for the bias-variance decomposition of  $\hat{h}_\ell$ . Goldenshluger and Lepski [16] developed a fully data-driven bandwidth selection method (GL method). The main idea of this method is based on an estimate of the bias term by looking at several estimators. In a similar fashion, Doumic et al. [14] and Reynaud-Bouret et al. [31] have used this method. To apply the GL method, we set for any  $\ell, \ell' \in H$ :

$$\hat{h}_{\ell, \ell'} := \frac{1}{M_T} \sum_{i=1}^{M_T} (K_\ell \star K_{\ell'}) (\gamma - \Gamma_i^1) = (K_\ell \star \hat{h}_{\ell'}) (\gamma).$$

Finally, the adaptive bandwidth and the estimator of  $h$  are selected as follows:

**Definition 4.** Given  $\epsilon > 0$  and setting  $\chi := (1 + \epsilon)(1 + \|K\|_1)$ , we define

$$\hat{\ell} := \operatorname{argmin}_{\ell \in H} \left\{ A(\ell) + \frac{\chi \|K\|_2}{\sqrt{M_T \ell}} \right\}, \quad (26)$$

where, for any  $\ell \in H$ ,

$$A(\ell) := \sup_{\ell' \in H} \left\{ \|\hat{h}_{\ell, \ell'} - \hat{h}_{\ell'}\|_2 - \frac{\chi \|K\|_2}{\sqrt{M_T \ell'}} \right\}_+, \quad (27)$$

Then, the estimator  $\hat{h}$  is given by

$$\hat{h} := \hat{h}_{\hat{\ell}}. \quad (28)$$

An inspection of the proof of Theorem 2 shows that the term  $A(\ell)$  provides a control for the bias  $\|h - K_\ell \star h\|_2$  up to the term  $\|K\|_1$  (see (45) and (47) in the proof of Theorem 2, section 4). Since  $A(\ell)$  depends only on  $\hat{h}_{\ell, \ell'}$  and  $\hat{h}_{\ell'}$ , the estimator  $\hat{h}$  can be computed in practice.

We shall now state an oracle inequality which highlights the bias-variance decomposition of the MISE of  $\hat{h}$ . We recall that the MISE of  $\hat{h}$  is the quantity  $\mathbb{E}[\|\hat{h} - h\|_2^2]$ .

**Theorem 2.** Let  $T > 0$  and assume that observations are taken on  $[0, T]$ . Let  $N_0$  be the number of mother cells at the beginning of divisions and  $M_T$  is the random number of divisions in  $[0, T]$ . Consider  $H$  a countable subset of  $\{\Delta^{-1} : \Delta = 1, \dots, \Delta_{\max}\}$  in which we choose the bandwidths and  $\Delta_{\max} = \lfloor \delta M_T \rfloor$  for some  $\delta > 0$ . Assume  $h \in L^\infty([0, 1])$  and let  $\hat{h}$  be a kernel estimator defined with the kernel  $K_{\hat{\ell}}$  where  $\hat{\ell}$  is chosen by the GL method. Define

$$\varrho(T)^{-1} = \begin{cases} \frac{e^{-RT + \log(RT)}}{1 - e^{-RT}}, & \text{if } N_0 = 1, \\ e^{-RT}, & \text{if } N_0 > 1. \end{cases} \quad (29)$$

For large  $T$ , the main term in  $\varrho(T)$  is  $e^{-RT}$  in any case. It is exactly the order of  $\varrho(T)$  for  $N_0 > 1$ . Then, given  $\epsilon > 0$

$$\mathbb{E}[\|\hat{h} - h\|_2^2] \leq C_1 \inf_{\ell \in H} \left\{ \|K_\ell \star h - h\|_2^2 + \frac{\|K\|_2^2}{\ell} \varrho(T)^{-1} \right\} + C_2 \varrho(T)^{-1}, \quad (30)$$

where  $C_1$  is a constant depending on  $N_0$ ,  $\|K\|_1$  and  $\epsilon$  and  $C_2$  is a constant depending on  $N_0$ ,  $\delta$ ,  $\epsilon$ ,  $\|K\|_1$ ,  $\|K\|_2$  and  $\|h\|_\infty$ .

The term  $\|K_\ell \star h - h\|_2^2$  is an approximation term,  $\frac{\|K\|_2^2}{\ell} \varrho(T)^{-1}$  is a variance term and the last term  $\varrho(T)^{-1}$  is asymptotically negligible. Hence the right hand side of the oracle inequality corresponds to a bias variance trade-off.

We now establish upper and lower bounds for the MISE. The lower bound is obtained by perturbation methods (Theorem 4) and is valid for any estimator  $\hat{h}_T$  of  $h$ , thus indicating the optimal convergence rate. The upper bound is obtained in Theorem 3 thanks to the key oracle inequality of Theorem 2.

For the rate of convergence, it is necessary to assume that the density  $h$  and the kernel function  $K$  satisfy some regularity conditions introduced in the following definitions.



**Definition 5.** Let  $\beta > 0$  and  $L > 0$ . The Hölder class of smoothness  $\beta$  and radius  $L$  is defined by

$$\mathcal{H}(\beta, L) = \left\{ f : f \text{ has } k = \lfloor \beta \rfloor \text{ derivatives and } \forall x, y \in \mathbb{R} \right. \\ \left. |f^{(k)}(y) - f^{(k)}(x)| \leq L|x - y|^{\beta - k} \right\}.$$

**Definition 6.** Let  $\beta^* > 0$ . An integrable function  $K : \mathbb{R} \rightarrow \mathbb{R}$  is a kernel of order  $\beta^*$  if

- $\int K(x)dx = 1$ ,
- $\int |x|^{\beta^*} |K(x)|dx < \infty$ ,
- For  $k = \lfloor \beta^* \rfloor$ ,  $\forall 1 \leq j \leq k$ ,  $\int x^j K(x)dx = 0$ .

Then, the following theorem gives the rate of convergence of the adaptive estimator  $\hat{h}$ .

**Theorem 3.** Let  $\beta^* > 0$  and  $K$  be a kernel of order  $\beta^*$ . Let  $\beta \in (0, \beta^*)$ . Let  $\hat{\ell}$  be the adaptive bandwidth defined in (26). Then, for any  $T > 0$ , the kernel estimator  $\hat{h}$  satisfies

$$\sup_{h \in \mathcal{H}(\beta, L)} \mathbb{E} \|\hat{h} - h\|_2^2 \leq C_3 \varrho(T)^{-\frac{2\beta}{2\beta+1}}, \quad (31)$$

where  $\varrho(T)^{-1}$  is defined in (29) and  $C_3$  is a constant depending on  $N_0$ ,  $\delta$ ,  $\epsilon$ ,  $\|K\|_1$ ,  $\|K\|_2$ ,  $\|h\|_\infty$ ,  $\beta$  and  $L$ .

We now establish a lower bound in Theorem 4.

**Theorem 4.** For any  $T > 0$ ,  $\beta > 0$  and  $L > 0$ . Assume that  $h \in \mathcal{H}(\beta, L)$ , then there exists a constant  $C_4 > 0$  such that for any estimator  $\hat{h}_T$  of  $h$

$$\sup_{h \in \mathcal{H}(\beta, L)} \mathbb{E} \|\hat{h}_T - h\|_2^2 \geq C_4 \exp \left( -\frac{2\beta}{2\beta+1} RT \right). \quad (32)$$

Contrary to the classical cases of nonparametric estimation (e.g. Tsybakov [36], ...), the number of observations  $M_T$  is a random variable that converges to  $+\infty$  when  $T \rightarrow +\infty$  which is one of the main difficulty here. From Theorem 3, when  $N_0 > 1$  the upper bound is in  $\exp \left( -\frac{2\beta}{2\beta+1} RT \right)$  which is the same rate as the lower bound. The rate of convergence  $\hat{h}$  is thus optimal. When  $N_0 = 1$ , the upper bound is in  $\exp \left( \frac{2\beta}{2\beta+1} (-RT + \log(RT)) \right)$  that differs with a logarithmic from the rate in the lower bound. The rate of convergence is thus slightly slower than in the case  $N_0 > 1$  and our estimator is optimal up to a logarithmic factor. Furthermore, Theorem 3 illustrates adaptive properties of our procedure: it achieves the rate  $\varrho(T)^{-\frac{2\beta}{2\beta+1}}$  over the Hölder class  $\mathcal{H}(\beta, L)$  as soon as  $\beta$  is smaller than  $\beta^*$ . So, it automatically adapts to the unknown smoothness of the signal to estimate.

### 3 Numerical simulations

#### 3.1 Numerical computation of $\hat{h}$

We use the **R** software to implement simulations with two original distributions of division kernel  $h$  and compare with their estimators. On the interval  $[0, 1]$ , the first distribution to

test is  $\text{Beta}(2, 2)$ .  $\text{Beta}(a, b)$  distributions on  $[0, 1]$  are characterized by their densities

$$h_{\text{Beta}(a,b)}(x) = \frac{x^{a-1}(1-x)^{b-1}}{\mathcal{B}(a,b)}.$$

where  $\mathcal{B}(a, b)$  is the renormalization constant.

Since  $h$  is symmetric, we only consider the distributions with  $a = b$ . Generally, asymmetric divisions correspond to  $a < 1$  and symmetric divisions with kernels concentrated around  $\frac{1}{2}$  correspond to  $a > 1$ . The smaller the parameter  $a$ , the more asymmetric the divisions. For the second density, we choose a Beta mixture distribution as

$$\frac{1}{2} \text{Beta}(2, 6) + \frac{1}{2} \text{Beta}(6, 2).$$

This choice gives us a bimodal density corresponding to very asymmetric divisions.

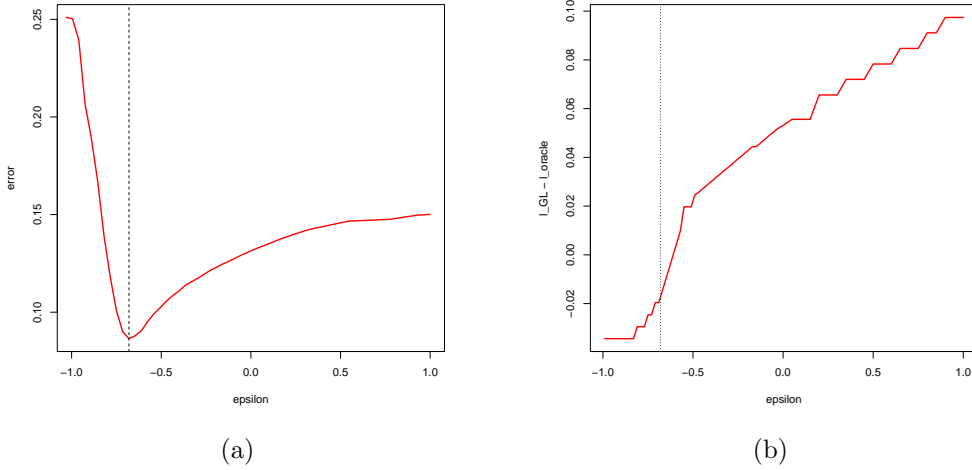


Figure 2: (a): MISE's as a function of  $\epsilon$ . (b):  $\hat{\ell} - \ell_{\text{oracle}}$  as a function of  $\epsilon$ . The dotted lines indicate the optimal value of  $\epsilon$  which is used in all simulations.

We estimate  $\hat{h}$  by using (19) and we take the classical Gaussian kernel  $K(x) = (2\pi)^{-1/2} \exp(-x^2/2)$ . For the choice of bandwidth, we apply the GL method with the family  $H = \{1, 2^{-1}, \dots, [\delta M_T]^{-1}\}$  for some  $\delta > 0$  small enough when  $M_T$  is large to reduce the time of numerical simulation. We have  $\|K\|_1 = 1$ ,  $\|K\|_2 = 2^{-1/2} \pi^{-1/4}$  and  $K_\ell \star K_{\ell'} = K_{\sqrt{\ell^2 + \ell'^2}}$ , hence it is not difficult to calculate in practice  $\hat{h}_{\ell, \ell'}$  as well as  $\hat{h}_{\ell'}$ . Finally, the value of  $\epsilon$  in  $\chi = (1 + \epsilon)(1 + \|K\|_1)$  is chosen to find an optimal value of the MISE. To do this, we implement a preliminary simulation to calibrate  $\epsilon$  in which we choose  $\epsilon > -1$  to ensure that  $1 + \epsilon > 0$ . We compute the MISE and  $\hat{\ell} - \ell_{\text{oracle}}$  as functions of  $\epsilon$  where  $\ell_{\text{oracle}} = \argmin_{\ell \in H} \mathbb{E}[\|\hat{h}_\ell - h\|_2^2]$  and  $h$  is the density of  $\text{Beta}(2, 2)$ . In Figure 2a, simulation results show that the risk has minimum value at  $\epsilon = -0.68$ . This value is not justified from a theoretical point of view. The theoretical choice  $\epsilon > 0$  (see Theorem 2) does not give bad results but this choice is too conservative for non-asymptotic practical purposes as often met in the literature (see Bertin et al. [9] for more details about the GL methodology). Moreover, following the discussion in Lacour and Massart [23] we investigate (see Figure 2b) the difference  $\hat{\ell} - \ell_{\text{oracle}}$  and observe some explosions close to  $\epsilon = -0.68$ . Consequently, we choose  $\epsilon = -0.68$  for all following simulations.

Figure 3 illustrates a reconstruction for the density of Beta(2, 2) and beta mixture  $\frac{1}{2}\text{Beta}(2, 6) + \frac{1}{2}\text{Beta}(6, 2)$  when  $T = 13$ . We choose here the division rate and the growth rate  $R = 0.5$  and  $\alpha = 0.35$  respectively. We compare the estimated densities when using the GL bandwidth with those estimated with the oracle bandwidth. The oracle bandwidth is found by assuming that we know the true density. Moreover, the GL estimators are compared with estimators using the cross-validation (CV) method and the rule of thumb (RoT). The CV bandwidth is defined as follows:

$$\ell_{CV} = \operatorname{argmin}_{\ell \in H} \left\{ \int \hat{h}_{\ell}^2(\gamma) d\gamma - \frac{2}{n} \sum_{i=1}^n \hat{h}_{\ell, -i}(\Gamma_i^1) \right\}$$

where  $\hat{h}_{\ell, -i}(\gamma) = \frac{1}{n-1} \sum_{j \neq i} K_{\ell}(\Gamma_j^1 - \gamma)$ . The RoT bandwidth can be calculated simply by using the formula  $\ell_{RoT} = 1.06\hat{\sigma}n^{-1/5}$  where  $\hat{\sigma}$  is the standard deviation of the sample  $(\Gamma_1^1, \dots, \Gamma_n^1)$ . More details about these methods can be found in Section 3.4 of Silverman [32] or Tsybakov [36].

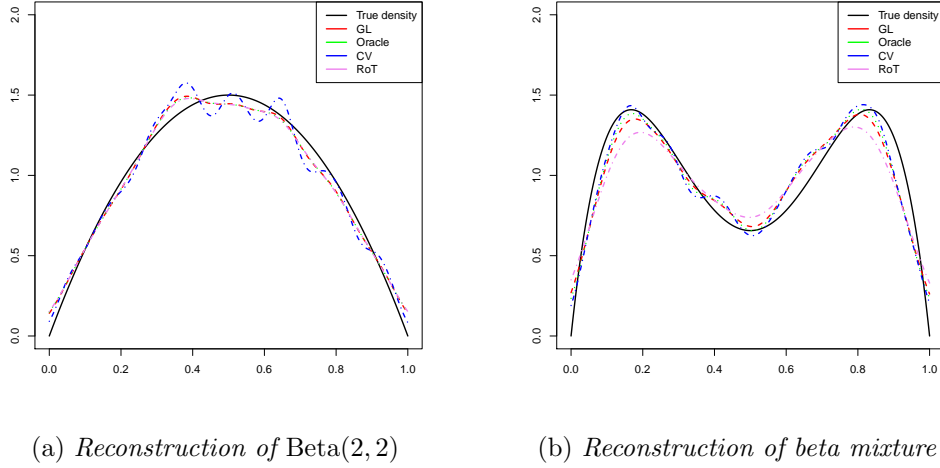


Figure 3: *Reconstruction of division kernels with  $T = 13$ .*

To estimate the MISE, we implement Monte-Carlo simulations with respect to  $T = 13, 17$  and  $20$ . The number of repetitions for each simulation is  $\mathcal{M} = 100$ . Then, we compute the mean of relative error  $\bar{e} = (1/\mathcal{M}) \sum_{i=1}^{\mathcal{M}} e_i$  and the standard deviation  $\sigma_e = \sqrt{(1/\mathcal{M}) \sum_{i=1}^{\mathcal{M}} (e_i - \bar{e})^2}$  where

$$e_i = \frac{\|\hat{h}^{(i)} - h\|_2}{\|h\|_2}, \quad i = 1, \dots, \mathcal{M}, \quad (33)$$

and  $\hat{h}^{(i)}$  denotes the estimator of  $h$  corresponding to  $i^{\text{th}}$  repetition.

The MISE's are computed for estimated densities using the GL bandwidth, the oracle bandwidth, the CV bandwidth and the RoT bandwidth. For a further comparison, in the reconstruction of Beta(2, 2), we compute the relative error in a parametric setting by comparing the true density  $h$  with the density of Beta( $\hat{a}, \hat{a}$ ) where  $\hat{a}$  is a Maximum Likelihood (ML) estimator  $a$ . The simulation results are displayed in Table 1 and Table 2. For the density of Beta mixture, we only compute the error with  $T = 13$  and  $T = 17$ . The boxplot in Figure 4 illustrates the MISE's in Table 1 when  $T = 17$ .

		GL	Oracle	CV	RoT	ML method
$T = 13$	$\bar{e}$	0.1001	0.0840	0.1009	0.0900	0.0610
	$\sigma_e$	0.0585	0.0481	0.0599	0.0577	0.0724
	$\hat{\bar{\ell}}$	0.0920	0.0845	0.0824	0.0727	
$T = 17$	$\bar{e}$	0.0458	0.0397	0.0459	0.0405	0.0166
	$\sigma_e$	0.0260	0.0230	0.0297	0.0237	0.0171
	$\hat{\bar{\ell}}$	0.0485	0.0497	0.0478	0.0470	
$T = 20$	$\bar{e}$	0.0261	0.0241	0.0262	0.0245	0.0088
	$\sigma_e$	0.0140	0.0114	0.0132	0.00121	0.0091
	$\hat{\bar{\ell}}$	0.0377	0.0359	0.0345	0.0354	

Table 1: Mean of relative error and its standard deviation for the reconstruction of Beta(2, 2).  $\hat{\bar{\ell}}$  is the average of bandwidths for  $M = 100$  samples.

		GL	Oracle	CV	RoT
$T = 13$	$\bar{e}$	0.1361	0.1245	0.1379	0.1686
	$\sigma_e$	0.0672	0.0562	0.0815	0.0537
	$\hat{\bar{\ell}}$	0.0618	0.0527	0.0522	0.0948
$T = 17$	$\bar{e}$	0.0539	0.0534	0.0550	0.0919
	$\sigma_e$	0.0180	0.0168	0.0168	0.00223
	$\hat{\bar{\ell}}$	0.0309	0.0272	0.0264	0.0590

Table 2: Mean of relative error and its standard deviation for the reconstruction of beta mixture  $\frac{1}{2} \text{Beta}(2, 6) + \frac{1}{2} \text{Beta}(6, 2)$ .

From Tables 1 and 2, we can note that the accuracy of the estimation of Beta(2, 2) and Beta mixture by the GL bandwidth increases for larger  $T$ . In Figure 5, we illustrate on a log-log scale the mean relative error and the rate of convergence versus time  $T$ . This shows that the error is close to the exponential rate predicted by the theory. Moreover, we can observe that the errors of Beta mixture are larger than those of Beta(2, 2) with the same  $T$  due to the complexity of its density. In both cases, the error estimated by using oracle bandwidth is always smaller. The GL error is slightly smaller than the CV error. The RoT error can show very good behavior but lacks of stability. Overall, we conclude that the GL method has a good behavior when compared to the cross validation method and rule-of-thumb. As usual, we also see that the ML errors are quite smaller than those of nonparametric approach but the magnitude of the mean  $\bar{e}$  remains similar.

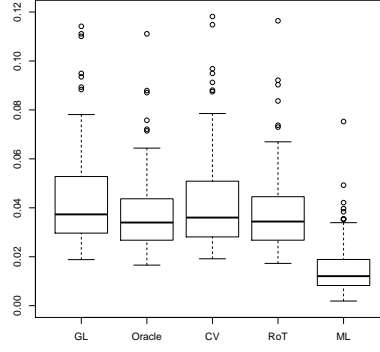


Figure 4: *Errors of estimated densities of Beta(2,2) when  $T = 17$ .*

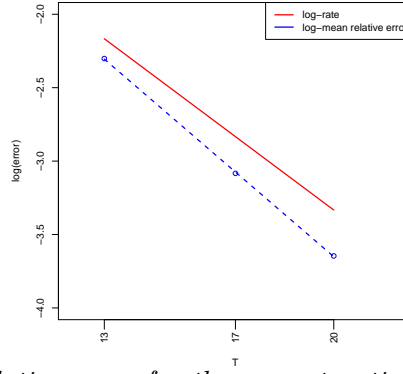


Figure 5: *The log-mean relative error for the reconstruction of Beta(2,2) compared to the log-rate (solid line) computed with  $\beta = 1$ .*

Since  $h$  is symmetric on  $[0, 1]$  with respect to  $\frac{1}{2}$ , the estimator  $\hat{h}$  can be improved and we can introduce

$$\tilde{h}(x) = \frac{1}{2} \left( \hat{h}(x) + \hat{h}(1-x) \right),$$

which is symmetric by construction and satisfies also (31). We compute the mean of relative error for the estimator  $\tilde{h}$  with the estimation of Beta(2,2) and Beta mixture. The results are displayed in Table 3. Compared with the error in Table 1 and 2, one can see as expected that the errors for the reconstruction of  $\tilde{h}$  are smaller. However, these errors are of the same order, indicating that the estimator  $\hat{h}$  had already good symmetric properties.

		GL	Oracle	CV	RoT
Beta(2,2)	$T = 13$	0.0785	0.0634	0.0762	0.0644
	$T = 17$	0.0356	0.0309	0.0356	0.0309
Beta mixture	$T = 13$	0.1117	0.0953	0.1030	0.1584
	$T = 17$	0.0450	0.0414	0.0417	0.0893

Table 3: *Mean of relative error for the reconstruction of  $\tilde{h}$ .*

### 3.2 Influence of the distribution on the mean age

For  $t \geq 0$ , recall the mean age defined in (13). To study the influence of the distribution on the mean age, we simulate  $n = 50$  trees with respect to  $t = 6, 6 + \Delta t, \dots, 24$  with  $\Delta t = 0.36$ . For each sample  $(\bar{x}_t^{(1)}, \dots, \bar{x}_t^{(n)})$ , we compute the average mean, the 1<sup>st</sup> ( $Q_{25}$ ) quartile and 3<sup>rd</sup> ( $Q_{75}$ ) quartile. Figure 6a and 6b show the simulation results corresponding to the density of Beta(2, 2) with  $\alpha = 0.45$  and  $R = 0.4$ . One can see that the average of mean age and the mean age converge to  $\frac{\alpha}{R} = 1.125$  for larger  $t$ . This agrees with the theoretical result proved in Section 2.2.

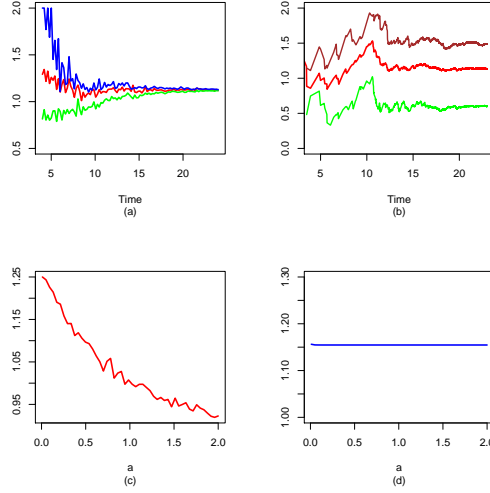


Figure 6: (a) Average mean, 1<sup>st</sup> and 3<sup>rd</sup> quartiles for the sample of means for 50 trees. (b) Average mean, 1<sup>st</sup> and 3<sup>rd</sup> quartiles for one tree. (c) Average of  $Q_{75} - Q_{25}$  with  $a \in [0, 2]$  at  $t = 12$ . (d) Mean age with  $a \in [0, 2]$  at  $t = 12$ .

Moreover,  $Q_{25}$  and  $Q_{75}$  vary when the parameter  $a$  changes. In Figure 6c, we draw a fitted curve of the average of  $(Q_{75} - Q_{25})$  when  $a$  varies from 0 to 2. As we mentioned in the introduction, if divisions are more asymmetric corresponding to small values of  $a$ , the toxicities concentrate on few cells, *i.e.* we have more older cells after the divisions. This explains the decreasing trend in the average of  $(Q_{75} - Q_{25})$ . Finally, Figure 6d displays the average of mean ages with respect to  $a$ . One can note that it does not change when we replace the kernel distribution, *e.g.* Beta(0.6, 0.6) instead of Beta(2, 2).

## 4 Proofs

### 4.1 Proof of Proposition 1

*ii)* The proof of *ii)* can be found easily in literature. Here we refer to [30], Section 5.3 for this proof.

*i)* Let us prove that  $\lim_{T \rightarrow +\infty} N_T = \lim_{j \rightarrow +\infty} N_{T_j} = +\infty$ . Since our model has only births and no death,  $(N_t)_{t \in [0, T]}$  is a non-decreasing process:  $N_{T_j} = N_0 + j$ . All the  $T_j$ 's are finite and  $\lim_{j \rightarrow +\infty} N_{T_j} = +\infty$  a.s. From *ii)*, we have  $\mathbb{E}[N_T] = N_0 e^{RT}$ . Hence, we deduce from the estimate  $\sup_{t \in [0, T]} \mathbb{E}[N_t] < +\infty$  for all  $T > 0$  that  $T_j \xrightarrow{j \rightarrow +\infty} +\infty$  a.s. Then we also have

$\lim_{T \rightarrow +\infty} N_T = +\infty$  a.s.

iii) Let  $p = e^{-RT}$ . When  $N_0 = 1$ ,  $N_T \sim \text{Geom}(p)$ . Then we have

$$\begin{aligned}\mathbb{E} \left[ \frac{1}{N_T} \right] &= \sum_{n=1}^{\infty} \frac{1}{n} \mathbb{P}(N_T = n) = \sum_{n=1}^{\infty} \frac{1}{n} p(1-p)^{n-1} \\ &= \frac{p}{1-p} \sum_{n=1}^{+\infty} \frac{(1-p)^n}{n} = -\frac{p}{1-p} \log(p).\end{aligned}$$

Replace  $p$  with  $e^{-RT}$ , we obtain (10).

When  $N_0 > 1$ ,  $N_T \sim \mathcal{NB}(N_0, p)$ . Hence, we have

$$\begin{aligned}\mathbb{E} \left[ \frac{1}{N_T} \right] &= \sum_{n=N_0}^{\infty} \frac{1}{n} \binom{n-1}{n-N_0} p^{N_0} (1-p)^{n-N_0} \\ &= \left( \frac{p}{1-p} \right)^{N_0} \sum_{n=N_0}^{\infty} \frac{1}{n} \binom{n-1}{n-N_0} (1-p)^n \\ &:= \left( \frac{p}{1-p} \right)^{N_0} f(1-p),\end{aligned}\tag{34}$$

where  $f(x) = \sum_{n=N_0}^{+\infty} \frac{1}{n} \binom{n-1}{n-N_0} x^n$ . We can differentiate  $f(x)$  by taking derivative under the sum. Then:

$$\begin{aligned}\frac{d}{dp} f(1-p) &= - \sum_{n=N_0}^{+\infty} \binom{n-1}{n-N_0} (1-p)^{n-1} \\ &= - \frac{(1-p)^{N_0-1}}{p^{N_0}} \sum_{n=N_0}^{+\infty} \binom{n-1}{n-N_0} p^{N_0} (1-p)^{n-N_0} = -\frac{1}{p} \left( \frac{1}{p} - 1 \right)^{N_0-1},\end{aligned}$$

since the sum is 1 (we recognize the negative binomial).

Hence,

$$\begin{aligned}\frac{d}{dp} f(1-p) &= -\frac{1}{p} \left[ \sum_{k=1}^{N_0-1} \binom{N_0-1}{k} \frac{1}{p^k} (-1)^{N_0-1-k} + (-1)^{N_0-1} \right] \\ &= (-1)^{N_0} \left[ \sum_{k=1}^{N_0-1} \binom{N_0-1}{k} \frac{(-1)^k}{p^{k+1}} + \frac{1}{p} \right].\end{aligned}\tag{35}$$

Integrating equation (35) and notice that  $f(0) = 0$ , we get

$$\begin{aligned}f(1-p) &= (-1)^{N_0} \left[ \sum_{k=1}^{N_0-1} \binom{N_0-1}{k} \frac{(-1)^k}{k} \left( -\frac{1}{p^k} \right) + \log(p) \right] \\ &= (-1)^{N_0-1} \left[ \sum_{k=1}^{N_0-1} \binom{N_0-1}{k} \frac{(-1)^k}{k} \frac{1}{p^k} + \log \left( \frac{1}{p} \right) \right].\end{aligned}\tag{36}$$

Combine (34),(36) and replace  $p$  with  $e^{-RT}$ , we get (11).

iv) We first prove the lower bound of (12). From (5), taking  $f_t(x) = 1$ , we have

$$N_T = N_0 + \int_0^T \int_{\mathcal{E}} \mathbb{1}_{\{i \leq N_{s-}\}} Q(ds, di, d\gamma). \quad (37)$$

Applying Itô formula for jump processes (see [20], Theorem 5.1 on p.67) to (37), we obtain

$$\begin{aligned} \frac{1}{N_T} &= \frac{1}{N_0} + \int_0^T \int_{\mathcal{E}} \left( \frac{1}{N_{s-} + 1} - \frac{1}{N_{s-}} \right) \mathbb{1}_{\{i \leq N_{s-}\}} Q(ds, di, d\gamma) \\ &= \frac{1}{N_0} - \int_0^T \int_{\mathcal{E}} \frac{1}{N_{s-} (N_{s-} + 1)} \mathbb{1}_{\{i \leq N_{s-}\}} Q(ds, di, d\gamma). \end{aligned}$$

Hence,

$$\mathbb{E} \left[ \frac{1}{N_T} \right] = \frac{1}{N_0} - \mathbb{E} \left[ \int_0^T \frac{1}{N_s (N_s + 1)} R N_s ds \right] = \frac{1}{N_0} - R \int_0^T \mathbb{E} \left[ \frac{1}{N_s + 1} \right] ds. \quad (38)$$

Since  $N_s \geq N_0$ , we have  $\frac{1}{N_s + 1} \leq \frac{1}{N_s}$ . Therefore, (38) implies that

$$\mathbb{E} \left[ \frac{1}{N_T} \right] \geq \frac{1}{N_0} - R \int_0^T \mathbb{E} \left[ \frac{1}{N_s} \right] ds. \quad (39)$$

By comparison of  $\mathbb{E} \left[ \frac{1}{N_T} \right]$  with the solutions of the ODE  $\frac{d}{dT} u(T) = -R u(T)$  with  $u(0) = 1/N_0$ , we finally obtain

$$\mathbb{E} \left[ \frac{1}{N_T} \right] \geq \frac{1}{N_0} e^{-RT}.$$

For the upper bound, notice that  $\mathbb{E} \left[ \frac{1}{N_T} \right] \leq \mathbb{E} \left[ \frac{1}{N_T - 1} \right]$  for  $N_0 > 1$ . Then we have

$$\begin{aligned} \mathbb{E} \left[ \frac{1}{N_T - 1} \right] &= \sum_{n=N_0}^{+\infty} \frac{1}{n-1} \binom{n-1}{n-N_0} p^{N_0} (1-p)^{n-N_0} \\ &= \sum_{n=N_0}^{+\infty} \frac{(n-2)!}{(n-N_0)!(N_0-1)!} p^{N_0} (1-p)^{n-N_0} \\ &= \frac{p}{N_0-1} \sum_{n=N_0}^{+\infty} \frac{(n-2)!}{(n-N_0)!(N_0-2)!} p^{N_0-1} (1-p)^{n-N_0} \\ &= \frac{p}{N_0-1} \sum_{m=N_0-1}^{+\infty} \frac{(m-1)!}{(m-(N_0-1))!((N_0-1)-1)!} p^{N_0-1} (1-p)^{m-(N_0-1)} \\ &= \frac{p}{N_0-1} = \frac{e^{-RT}}{N_0-1}, \end{aligned}$$

by changing the index in the sum ( $m = n - 1$ ) and by recognizing the negative binomial with parameter  $(N_0 - 1, p)$ . Hence, we conclude that for  $N_0 > 1$

$$\frac{e^{-RT}}{N_0} \leq \mathbb{E} \left[ \frac{1}{N_T} \right] \leq \frac{e^{-RT}}{N_0 - 1}.$$

This ends the proof of Proposition 1.



## 4.2 Proof of Lemma 1

By symmetry of  $h$  with respect to  $1/2$ , we have:

$$\begin{aligned} Y_t &= Y_0 + \int_0^t \left( \alpha + 2R \int_0^1 (\gamma Y_s - Y_s) h(\gamma) d\gamma \right) ds + U_t \\ &= Y_0 + \int_0^t \left( \alpha - 2RY_s \int_0^1 \gamma h(\gamma) d\gamma \right) ds + U_t \\ &= Y_0 + \int_0^t (\alpha - RY_s) ds + U_t. \end{aligned}$$

where  $U_t$  is a square-integrable martingale.

Let  $\tilde{Y}_t = Y_t e^{Rt}$ ,  $\tilde{Y}_0 = Y_0$ . By Itô formula, we get

$$\tilde{Y}_t = \tilde{Y}_0 + \frac{\alpha}{R} (e^{Rt} - 1) + \int_0^t e^{Rs} dU_s.$$

Replacing  $\tilde{Y}_t$  by  $Y_t e^{Rt}$ , we obtain

$$Y_t = \left( Y_0 - \frac{\alpha}{R} \right) e^{-Rt} + \frac{\alpha}{R} + \int_0^t e^{-R(t-s)} dU_s.$$

We end the proof by taking the expectation and the limit as  $t \rightarrow +\infty$  of  $Y_t$  to obtain (16) and (17).

## 4.3 Proof of Theorem 1

We will show that the process  $Y$  satisfies ergodicity and integrability assumptions in Bansaye et al. [5] (see (H1) - (H4), Section 4). More precisely:

1.  $\mathbb{E}[Y_t] < +\infty$  for all  $t \geq 0$ .
2. There exists  $\varpi < R$  and  $c > 0$  such that  $\mathbb{E}[Y_t^2] < ce^{\varpi t}$  for all  $t \geq 0$ .

From (16) we note that  $\mathbb{E}[Y_t] < +\infty$  for all  $t \geq 0$ . To prove the second point, from (14) we have

$$\begin{aligned} \mathbb{E}[Y_t^2] &= \mathbb{E} \left[ Y_0^2 + \int_0^t \left( 2\alpha Y_s + 2R \int_0^1 (\gamma^2 Y_s^2 - Y_s^2) h(\gamma) d\gamma \right) ds \right] \\ &= Y_0^2 + 2\alpha \int_0^t \mathbb{E}[Y_s] ds - 2\theta R \int_0^t \mathbb{E}[Y_s^2] ds, \end{aligned} \tag{40}$$

with  $\theta = \int_0^1 (1 - \gamma^2) h(\gamma) d\gamma$  and  $0 < \theta < 1$ .

Substituting  $\mathbb{E}[Y_t] = (Y_0 - \alpha/R)e^{-Rt} + \alpha/R$  into (40), we see that  $\mathbb{E}(Y_t^2)$  solves the following equation:

$$\frac{d\mathbb{E}[Y_t^2]}{dt} = -2\theta R \mathbb{E}[Y_t^2] + \left( 2\alpha Y_0 - \frac{2\alpha^2}{R} \right) e^{-Rt} + \frac{2\alpha^2}{R}. \tag{41}$$

The solution of the equation (41) is:

$$\mathbb{E}[Y_t^2] = e^{-2\theta Rt} \left[ Y_0^2 + \int_0^t e^{2\theta Rs} \left( (2\alpha Y_0 - \frac{2\alpha^2}{R})e^{-Rs} + \frac{2\alpha^2}{R} \right) ds \right]. \quad (42)$$

Hence, if  $\theta = \frac{1}{2}$ , we have

$$\begin{aligned} \mathbb{E}[Y_t^2] &= Y_0^2 e^{-Rt} + (2\alpha Y_0 - \frac{2\alpha^2}{R}) t e^{-Rt} + \frac{2\alpha^2}{R^2} (1 - e^{-Rt}) \\ &\leq Y_0^2 e^{-Rt} + (2\alpha Y_0 - \frac{2\alpha^2}{R}) e^{-(R-\theta)t} + \frac{2\alpha^2}{R^2} \\ &\leq \left( Y_0^2 + 2\alpha Y_0 + \frac{2\alpha^2}{R} + \frac{2\alpha^2}{R^2} \right) e^{(0 \vee (\theta - R))t} = c_1 e^{\varpi t}, \end{aligned}$$

with  $\varpi = 0 \vee (\theta - R) := \max(0, \theta - R)$ .

If  $\theta \neq \frac{1}{2}$ ,

$$\begin{aligned} \mathbb{E}[Y_t^2] &= e^{-2\theta Rt} \left[ Y_0^2 + (2\alpha Y_0 - \frac{2\alpha^2}{R}) \int_0^t e^{(2\theta-1)Rs} ds + \frac{2\alpha^2}{R} \int_0^t e^{2\theta Rs} ds \right] \\ &= Y_0^2 e^{-2\theta Rt} + (2\alpha Y_0 - \frac{2\alpha^2}{R}) \frac{1}{(2\theta-1)R} (e^{-Rt} - e^{-2\theta Rt}) \\ &\quad + \frac{\alpha^2}{\theta R^2} (1 - e^{-2\theta Rt}) \\ &\leq \left( Y_0^2 + (2\alpha Y_0 + \frac{2\alpha^2}{R}) \frac{1}{|2\theta-1|R} + \frac{\alpha^2}{\theta R^2} \right) = c_2. \end{aligned}$$

Thus, if we set  $c = \max(c_1, c_2)$  then  $\mathbb{E}[Y_t^2] < c e^{\varpi t}$  for all  $t \geq 0$ .

The infinitesimal generator  $A$  of  $Y$  is defined for  $\mathcal{C}^1$  test functions as

$$Af(x) = \alpha f'(x) + 2R \int_0^1 (f(\gamma x) - f(x)) h(\gamma) d\gamma.$$

For  $V(x) = x$  and  $f(x) = x + 1$ , we have

$$AV(x) = \alpha - Rx \leq -\frac{R}{2} f(x) + \left( \alpha + \frac{R}{2} \right) \mathbb{1}_{\{x \leq \frac{2\alpha}{R} + 1\}}.$$

Hence, by Theorem 5.3 of Meyn and Tweedie [26], there exists  $\pi \in \mathcal{M}_F(\mathbb{R}_+)$  such that  $\lim_{t \rightarrow +\infty} E[Y_t] = \langle \pi, f \rangle = \frac{\alpha}{R}$ . Finally, applying Theorem 4.2 of [5], we obtain the result

$$\lim_{t \rightarrow +\infty} \frac{\langle Z_t, f \rangle}{N_t} = \langle \pi, f \rangle = \frac{\alpha}{R}.$$

#### 4.4 Proof of Proposition 2

To prove (20), let us remark that the number of random divisions  $M_T$  is independent of  $(\Gamma_i^1)_{i \in \mathbb{N}^*}$ , because the division rate  $R$  is constant and because of the construction of our

stochastic process. Therefore, we have

$$\begin{aligned}\mathbb{E}[\hat{h}_\ell|M_T] &= \mathbb{E}\left[\frac{1}{M_T} \sum_{i=1}^{M_T} K_\ell(\gamma - \Gamma_i^1)|M_T\right] = \frac{M_T \mathbb{E}[K_\ell(\gamma - \Gamma_1^1)]}{M_T} \\ &= \mathbb{E}[K_\ell(\gamma - \Gamma_1^1)] = K_\ell \star h(\gamma),\end{aligned}$$

and  $\mathbb{E}[\hat{h}_\ell] = \mathbb{E}[\mathbb{E}[\hat{h}_\ell|M_T]] = K_\ell \star h(\gamma)$ . By similar calculations as (20), we obtain (21) and (22).

To prove *ii*), by the Strong Law of Large Numbers, we have

$$\frac{1}{n} \sum_{i=1}^n K_\ell(\gamma - \Gamma_i^1) \xrightarrow{\text{a.s.}} \mathbb{E}[K_\ell(\gamma - \Gamma_1^1)] \quad \text{as } n \rightarrow +\infty.$$

From (6), we have  $\lim_{T \rightarrow +\infty} N_T = +\infty$  (a.s.). Since  $M_T = N_T - N_0$  and  $N_0$  is deterministic, this yields

$$\frac{1}{M_T} \sum_{i=1}^{M_T} K_\ell(\gamma - \Gamma_i^1) \xrightarrow{\text{a.s.}} \mathbb{E}[K_\ell(\gamma - \Gamma_1^1)] = K_\ell \star h(\gamma).$$

This ends the proof of Proposition 2.

## 4.5 Proof of Proposition 3

We have

$$\mathbb{E}[\|\hat{h}_\ell - h\|_2|M_T] \leq \|h - K_\ell \star h\|_2 + \mathbb{E}[\|\hat{h}_\ell - \mathbb{E}[\hat{h}_\ell]\|_2|M_T].$$

For the variance term, using that  $\mathbb{E}[\hat{h}_\ell(\gamma)] = \mathbb{E}[\hat{h}_\ell(\gamma)|M_T]$

$$\begin{aligned}\mathbb{E}[\|\hat{h}_\ell - \mathbb{E}[\hat{h}_\ell]\|_2^2|M_T] &= \mathbb{E}\left[\int_{\mathbb{R}} |\hat{h}_\ell(\gamma) - \mathbb{E}[\hat{h}_\ell(\gamma)]|^2 d\gamma|M_T\right] \\ &= \int_{\mathbb{R}} \mathbb{V}ar[\hat{h}_\ell(\gamma)|M_T] d\gamma \\ &= \frac{1}{M_T} \int_{\mathbb{R}} \mathbb{V}ar[K_\ell(\gamma - \Gamma_1^1)] d\gamma \\ &\leq \frac{1}{M_T} \int_{\mathbb{R}} \mathbb{E}[K_\ell^2(\gamma - \Gamma_1^1)] d\gamma\end{aligned}$$

By Fubini's theorem, we get

$$\begin{aligned}\int_{\mathbb{R}} \mathbb{E}[K_\ell^2(\gamma - \Gamma_1^1)] d\gamma &= \int_{\mathbb{R}} \int_{\mathbb{R}} K_\ell^2(\gamma - u) h(u) du d\gamma \\ &= \int_{\mathbb{R}} h(u) \left( \int_{\mathbb{R}} K_\ell^2(\gamma - u) d\gamma \right) du \\ &= \|K_\ell\|_2^2 \int_{\mathbb{R}} h(u) du = \frac{\|K\|_2^2}{\ell}.\end{aligned}$$

Then we have

$$\mathbb{E}[\|\hat{h}_\ell - \mathbb{E}[\hat{h}_\ell]\|_2^2|M_T] \leq \frac{\|K\|_2^2}{M_T \ell}. \quad (43)$$

Hence, applying Cauchy-Schwarz's inequality, we obtain (24). This ends the proof of Proposition 3.

## 4.6 Proof of Theorem 2

This proof is inspired by the proof of Doumic et al. [14]. However, our problem here is that the number of observations  $M_T$  is random. To overcome this difficulty, we work conditionally to  $M_T$  to get concentration inequalities.

Hereafter, we refer  $\int f$  to  $\int_{\mathbb{R}} f$  and since the support of  $h$  is  $(0, 1)$ , we can write  $\int h(\gamma)d\gamma$  instead of  $\int_0^1 h(\gamma)d\gamma$ . Recall that

$$A(\ell) := \sup_{\ell' \in H} \left\{ \|\hat{h}_{\ell, \ell'} - \hat{h}_{\ell'}\|_2 - \frac{\chi \|K\|_2}{\sqrt{M_T \ell'}} \right\}_+.$$

Then, for any  $\ell \in H$ , we have

$$\|\hat{h} - h\|_2 \leq A_1 + A_2 + A_3,$$

where

$$\begin{aligned} A_1 &:= \|\hat{h} - \hat{h}_{\hat{\ell}, \ell}\|_2 \leq A(\ell) + \frac{\chi \|K\|_2}{\sqrt{M_T \hat{\ell}}}, \\ A_2 &:= \|\hat{h}_{\hat{\ell}, \ell} - \hat{h}_{\ell}\|_2 \leq A(\hat{\ell}) + \frac{\chi \|K\|_2}{\sqrt{M_T \hat{\ell}}}, \\ A_3 &:= \|\hat{h}_{\ell} - h\|_2. \end{aligned}$$

By definition of  $\hat{\ell}$ , we have

$$A_1 + A_2 \leq 2A(\ell) + 2\frac{\chi \|K\|_2}{\sqrt{M_T \hat{\ell}}}, \quad (44)$$

and

$$\begin{aligned} A(\ell) &\leq \sup_{\ell' \in H} \left\{ \|(\hat{h}_{\ell, \ell'} - \mathbb{E}[\hat{h}_{\ell, \ell'}]) - (\hat{h}_{\ell'} - \mathbb{E}[\hat{h}_{\ell'}])\|_2 \right. \\ &\quad \left. + \|\mathbb{E}[\hat{h}_{\ell, \ell'}] - \mathbb{E}[\hat{h}_{\ell'}]\|_2 - \frac{\chi \|K\|_2}{\sqrt{M_T \ell'}} \right\}_+ \\ &\leq \xi_T(\ell) + \sup_{\ell' \in H} \left\{ \|\mathbb{E}[\hat{h}_{\ell, \ell'}] - \mathbb{E}[\hat{h}_{\ell'}]\|_2 \right\}, \end{aligned} \quad (45)$$

where

$$\xi_T(\ell) = \sup_{\ell' \in H} \left\{ \|(\hat{h}_{\ell, \ell'} - \mathbb{E}[\hat{h}_{\ell, \ell'}]) - (\hat{h}_{\ell'} - \mathbb{E}[\hat{h}_{\ell'}])\|_2 - \frac{\chi \|K\|_2}{\sqrt{M_T \ell'}} \right\}_+. \quad (46)$$

For the term  $\sup_{\ell' \in H} \left\{ \|\mathbb{E}[\hat{h}_{\ell, \ell'}] - \mathbb{E}[\hat{h}_{\ell'}]\|_2 \right\}$ , we have

$$\begin{aligned} \mathbb{E}[\hat{h}_{\ell, \ell'}] - \mathbb{E}[\hat{h}_{\ell'}] &= \int (K_{\ell} \star K_{\ell'})(\gamma - u)h(u)du - \int K_{\ell'}(\gamma - v)h(v)dv \\ &= \int \int K_{\ell}(\gamma - u - t)K_{\ell'}(t)h(u)dtdu - \int K_{\ell'}(\gamma - v)h(v)dv \\ &= \int \int K_{\ell}(v - u)K_{\ell'}(\gamma - v)h(u)dudv - \int K_{\ell'}(\gamma - v)h(v)dv \\ &= \int K_{\ell'}(\gamma - v) \left( \int K_{\ell}(v - u)h(u)du - h(v) \right) dv \\ &= \int K_{\ell'}(\gamma - v) (K_{\ell} \star h(v) - h(v)) dv. \end{aligned}$$

Hence, we derive

$$\|\mathbb{E}[\hat{h}_{\ell,\ell'}] - \mathbb{E}[\hat{h}_{\ell'}]\|_2 = \|K_{\ell'} \star (K_{\ell} \star h - h)\|_2 \leq \|K\|_1 \|K_{\ell} \star h - h\|_2, \quad (47)$$

where the right hand side does not depend on  $\ell'$  allowing us to take  $\sup_{\ell' \in H}$  in the left hand side.

Thus, (44), (46) and (47) give

$$A_1 + A_2 \leq 2\xi_T(\ell) + 2\|K\|_1 \|K_{\ell} \star h - h\|_2 + 2\frac{\chi\|K\|_2}{\sqrt{M_T\ell}}.$$

Then,

$$\mathbb{E}[(A_1 + A_2)^2] \leq 12\mathbb{E}[\xi_T^2(\ell)] + 12\|K\|_1^2 \|K_{\ell} \star h - h\|_2^2 + 12\frac{\chi^2\|K\|_2^2}{\ell} \mathbb{E}\left[\frac{1}{M_T}\right]. \quad (48)$$

For the term  $A_3$ , we have from (43)

$$\begin{aligned} \mathbb{E}[A_3^2] &= \|\mathbb{E}[\hat{h}_{\ell}] - h\|_2^2 + \mathbb{E}[\|\hat{h}_{\ell} - \mathbb{E}[\hat{h}_{\ell}]\|_2^2] \\ &\leq \|K_{\ell} \star h - h\|_2^2 + \frac{\|K\|_2^2}{\ell} \mathbb{E}\left[\frac{1}{M_T}\right]. \end{aligned}$$

Finally, replacing  $\chi$  by  $(1 + \epsilon)(1 + \|K\|_1)$ , we have for any  $\ell \in H$

$$\begin{aligned} \mathbb{E}[\|\hat{h} - h\|_2^2] &\leq 2\mathbb{E}[(A_1 + A_2)^2] + 2\mathbb{E}[A_3^2] \\ &\leq 24\mathbb{E}[\xi_T^2(\ell)] + 2\left(1 + 12\|K\|_1^2\right) \|K_{\ell} \star h - h\|_2^2 \\ &\quad + 2\left(1 + 12(1 + \epsilon)^2(1 + \|K\|_1)^2\right) \frac{\|K\|_2^2}{\ell} \mathbb{E}\left[\frac{1}{M_T}\right] \\ &\leq 24\mathbb{E}[\xi_T^2(\ell)] + C_1 \left( \|K_{\ell} \star h - h\|_2^2 + \frac{\|K\|_2^2}{\ell} \mathbb{E}\left[\frac{1}{M_T}\right] \right), \end{aligned} \quad (49)$$

with  $C_1$  a constant depending on  $\epsilon$  and  $\|K\|_1$ .

It remains to deal with the term  $\mathbb{E}[\xi_T^2(\ell)]$  where  $\xi_T(\ell)$  is defined in (46),

$$\begin{aligned} \xi_T(\ell) &\leq \sup_{\ell' \in H} \left\{ \|\hat{h}_{\ell,\ell'} - \mathbb{E}[\hat{h}_{\ell,\ell'}]\|_2 + \|\hat{h}_{\ell'} - \mathbb{E}[\hat{h}_{\ell'}]\|_2 - \frac{\chi\|K\|_2}{\sqrt{M_T\ell'}} \right\}_+ \\ &\leq \sup_{\ell' \in H} \left\{ \|\hat{h}_{\ell'} - \mathbb{E}[\hat{h}_{\ell'}]\|_2 \|K\|_1 + \|\hat{h}_{\ell'} - \mathbb{E}[\hat{h}_{\ell'}]\|_2 - \frac{\chi\|K\|_2}{\sqrt{M_T\ell'}} \right\}_+ \\ &\leq \sup_{\ell' \in H} \left\{ (1 + \|K\|_1) \|\hat{h}_{\ell'} - \mathbb{E}[\hat{h}_{\ell'}]\|_2 - \frac{(1 + \epsilon)(1 + \|K\|_1)\|K\|_2}{\sqrt{M_T\ell'}} \right\}_+ \\ &\leq (1 + \|K\|_1) S_T, \end{aligned}$$

where

$$S_T := \sup_{\ell \in H} \left\{ \|\hat{h}_{\ell} - \mathbb{E}[\hat{h}_{\ell}]\|_2 - \frac{(1 + \epsilon)\|K\|_2}{\sqrt{M_T\ell}} \right\}_+.$$

Hence,

$$\mathbb{E}[\xi_T^2(\ell)] \leq (1 + \|K\|_1)^2 \mathbb{E}[\mathbb{E}[S_T^2 | M_T]]. \quad (50)$$

If we show that

$$\mathbb{E}[S_T^2 | M_T = n] \leq C_* \frac{1}{n}, \quad (51)$$

then

$$\mathbb{E}[\xi_T^2(\ell)] \leq C_*(1 + \|K\|_1)^2 \mathbb{E}\left[\frac{1}{M_T}\right] \quad (52)$$

where  $C_*$  is a constant.

Let us establish (51). When  $M_T = n$ ,  $\forall n \in \mathbb{N}^*$ , we set

$$\mathbb{E}[\Sigma_n^2] = \mathbb{E}[S_T^2 | M_T = n]$$

where

$$\Sigma_n := \sup_{\ell \in H} \left\{ \|Z_\ell\|_2 - \frac{(1 + \epsilon)\|K\|_2}{\sqrt{n\ell}} \right\}_+,$$

with

$$Z_\ell = \hat{h}_\ell - \mathbb{E}[\hat{h}_\ell] = \frac{1}{n} \sum_{i=1}^n K_\ell(\gamma - \Gamma_i^1) - \mathbb{E}[K_\ell(\gamma - \Gamma_i^1)].$$

Then,

$$\begin{aligned} \mathbb{E}[\Sigma_n^2] &= \mathbb{E}\left[\sup_{\ell \in H} \left\{ \|Z_\ell\|_2 - \frac{(1 + \epsilon)\|K\|_2}{\sqrt{n\ell}} \right\}_+^2\right] \\ &\leq \int_0^{+\infty} \mathbb{P}\left[\sup_{\ell \in H} \left\{ \|Z_\ell\|_2 - \frac{(1 + \epsilon)\|K\|_2}{\sqrt{n\ell}} \right\}_+^2 \geq x\right] dx \\ &\leq \sum_{\ell \in H} \int_0^{+\infty} \mathbb{P}\left[\left\{ \|Z_\ell\|_2 - \frac{(1 + \epsilon)\|K\|_2}{\sqrt{n\ell}} \right\}_+^2 \geq x\right] dx. \end{aligned}$$

We bound this with Talagrand's inequality.

Let  $\mathcal{A}$  be a countable dense subset of the unit ball of  $\mathbb{L}_2([0, 1])$ . We express the norm  $\|Z_\ell\|_2$  as

$$\begin{aligned} \|Z_\ell\|_2 &= \sup_{a \in \mathcal{A}} \int a(\gamma) Z_\ell(\gamma) d\gamma \\ &= \sup_{a \in \mathcal{A}} \sum_{i=1}^n \int a(\gamma) \frac{1}{n} \left( K_\ell(\gamma - \Gamma_i^1) - \mathbb{E}[K_\ell(\gamma - \Gamma_i^1)] \right) d\gamma. \end{aligned}$$

Let

$$V_{i,\Gamma} = \int a(\gamma) \frac{1}{n} \left( K_\ell(\gamma - \Gamma_i^1) - \mathbb{E}[K_\ell(\gamma - \Gamma_i^1)] \right) d\gamma.$$

Then  $V_{i,\Gamma}$ ,  $i = 1, \dots, n$  is a sequence of i.i.d random variables with zero mean. Thus, we can apply Talagrand's inequality (see [25, p. 170]) to  $\|Z_\ell\|_2 = \sup_{a \in \mathcal{A}} \sum_{i=1}^n V_{i,\Gamma}$ . For all  $\eta, x > 0$ , one has

$$\mathbb{P}\left(\|Z_\ell\|_2 \geq (1 + \eta)\mathbb{E}[\|Z_\ell\|_2] + \sqrt{2\nu x} + c(\eta)bx\right) \leq e^{-x},$$

where  $c(\eta) = 1/3 + \eta^{-1}$ ,

$$\nu = \frac{1}{n} \sup_{a \in \mathcal{A}} \mathbb{E}\left[\left(\int a(\gamma) \left(K_\ell(\gamma - \Gamma_1^1) - \mathbb{E}[K_\ell(\gamma - \Gamma_1^1)]\right) d\gamma\right)^2\right],$$

and,

$$b = \frac{1}{n} \sup_{y \in (0,1), a \in \mathcal{A}} \int a(\gamma) \left( K_\ell(\gamma - y) - \mathbb{E}[K_\ell(\gamma - \Gamma_1^1)] \right) d\gamma.$$

Next, we calculate the terms  $\mathbb{E}[\|Z_\ell\|_2]$ ,  $\nu$  and  $b$ . Applying Cauchy - Schwarz's inequality and using independence of variables, we get

$$\begin{aligned} \mathbb{E}[\|Z_\ell\|_2] &\leq \left( \mathbb{E}[\|Z_\ell\|_2^2] \right)^{1/2} \\ &\leq \left( \mathbb{E} \left[ \int \left( \frac{1}{n} \sum_{i=1}^n K_\ell(\gamma - \Gamma_i^1) - \mathbb{E}[K_\ell(\gamma - \Gamma_i^1)] \right)^2 d\gamma \right] \right)^{1/2} \\ &= \frac{1}{n} \left( \int \mathbb{E} \left[ \left( \sum_{i=1}^n K_\ell(\gamma - \Gamma_i^1) - \mathbb{E}[K_\ell(\gamma - \Gamma_i^1)] \right)^2 \right] d\gamma \right)^{1/2} \\ &= \frac{1}{n} \left( \int \sum_{i=1}^n \mathbb{E} \left[ \left( K_\ell(\gamma - \Gamma_i^1) - \mathbb{E}[K_\ell(\gamma - \Gamma_i^1)] \right)^2 \right] d\gamma \right)^{1/2} \\ &\leq \frac{1}{n} \left( n \int \mathbb{E} \left[ K_\ell(\gamma - \Gamma_1^1)^2 \right] d\gamma \right)^{1/2} = \frac{\|K\|_2}{\sqrt{n\ell}}. \end{aligned}$$

For the term  $\nu$ , we have

$$\begin{aligned} \nu &\leq \frac{1}{n} \sup_{a \in \mathcal{A}} \mathbb{E} \left[ \left( \int a(\gamma) K_\ell(\gamma - \Gamma_1^1) d\gamma \right)^2 \right] \\ &\leq \frac{1}{n} \sup_{a \in \mathcal{A}} \mathbb{E} \left[ \int |K_\ell(\gamma - \Gamma_1^1)| d\gamma \times \int a^2(\gamma) |K_\ell(\gamma - \Gamma_1^1)| d\gamma \right] \\ &\leq \frac{\|K\|_1}{n} \sup_{a \in \mathcal{A}} \mathbb{E} \left[ \int a^2(\gamma) |K_\ell(\gamma - \Gamma_1^1)| d\gamma \right] \\ &\leq \frac{\|K\|_1}{n} \sup_{a \in \mathcal{A}} \int a^2(\gamma) \mathbb{E} \left[ |K_\ell(\gamma - \Gamma_1^1)| \right] d\gamma \\ &\leq \frac{\|K\|_1}{n} \sup_{a \in \mathcal{A}} \int \int a^2(\gamma) |K_\ell(\gamma - u)| h(u) du d\gamma \\ &\leq \frac{\|h\|_\infty \|K\|_1^2}{n}. \end{aligned}$$

For the term  $b$ , we have

$$\begin{aligned} b &= \frac{1}{n} \sup_{y \in (0,1)} \|K_\ell(\cdot - y) - \mathbb{E}[K_\ell(\cdot - \Gamma_1^1)]\|_2 \\ &\leq \frac{1}{n} \left( \sup_{y \in (0,1)} \|K_\ell(\cdot - y)\|_2 + \left( \mathbb{E} \left[ \int K_\ell^2(\gamma - \Gamma_1^1) d\gamma \right] \right)^{1/2} \right) \leq \frac{2\|K\|_2}{n\sqrt{\ell}}. \end{aligned}$$

So, for all  $\eta, x > 0$ , we have

$$\mathbb{P} \left( \|Z_\ell\|_2 \geq (1 + \eta) \frac{\|K\|_2}{\sqrt{n\ell}} + \|h\|_\infty^{1/2} \|K\|_1 \sqrt{\frac{2x}{n}} + 2c(\eta) \frac{\|K\|_2}{n\sqrt{\ell}} \right) \leq e^{-x}.$$

Let  $W_\ell$  be some strictly positive weights, we apply the previous inequality to  $x = W_\ell + u$  for  $u > 0$ . We have

$$\begin{aligned} \mathbb{P} \left( \|Z_\ell\|_2 \geq (1+\eta) \frac{\|K\|_2}{\sqrt{n\ell}} + \|h\|_\infty^{1/2} \|K\|_1 \sqrt{\frac{W_\ell}{n}} + 2c(\eta) \frac{\|K\|_2 W_\ell}{n\sqrt{\ell}} \right. \\ \left. + \|h\|_\infty^{1/2} \|K\|_1 \sqrt{\frac{u}{n}} + 2c(\eta) \frac{\|K\|_2 u}{n\sqrt{\ell}} \right) \leq e^{-W_\ell - u}. \end{aligned}$$

If we set

$$\Psi_\ell = (1+\eta) \frac{\|K\|_2}{\sqrt{n\ell}} + \|h\|_\infty^{1/2} \|K\|_1 \sqrt{\frac{W_\ell}{n}} + 2c(\eta) \frac{\|K\|_2 W_\ell}{n\sqrt{\ell}},$$

then,

$$\mathbb{P} \left( \|Z_\ell\|_2 - \Psi_\ell \geq \|h\|_\infty^{1/2} \|K\|_1 \sqrt{\frac{u}{n}} + 2c(\eta) \frac{\|K\|_2 u}{n\sqrt{\ell}} \right) \leq e^{-W_\ell - u}.$$

Let

$$\Lambda = \mathbb{E} \left[ \sup_{\ell \in H} (\|Z_\ell\|_2 - \Psi_\ell)_+^2 \right] = \int_0^{+\infty} \mathbb{P} \left[ \sup_{\ell \in H} (\|Z_\ell\|_2 - \Psi_\ell)_+^2 \geq x \right] dx.$$

An upper bound of  $\Lambda$  is given by

$$\Lambda \leq \sum_{\ell \in H} \int_0^{+\infty} \mathbb{P} \left[ (\|Z_\ell\|_2 - \Psi_\ell)_+^2 \geq x \right] dx.$$

Let us take  $u$  such that

$$x = f(u)^2 = \left( \|h\|_\infty^{1/2} \|K\|_1 \sqrt{\frac{u}{n}} + 2c(\eta) \frac{\|K\|_2 u}{n\sqrt{\ell}} \right)^2.$$

So,

$$dx = 2f(u) \left( \|h\|_\infty^{1/2} \|K\|_1 \frac{1}{2\sqrt{nu}} + 2c(\eta) \frac{\|K\|_2}{n\sqrt{\ell}} \right) du.$$

Hence,

$$\begin{aligned} \Lambda &\leq \sum_{\ell \in H} \int_0^{+\infty} e^{-W_\ell - u} 2f(u) \left( \|h\|_\infty^{1/2} \|K\|_1 \frac{1}{2\sqrt{nu}} + 2c(\eta) \frac{\|K\|_2}{n\sqrt{\ell}} \right) du \\ &\leq \sum_{\ell \in H} \int_0^{+\infty} e^{-W_\ell - u} 2f(u) \left( \|h\|_\infty^{1/2} \|K\|_1 \sqrt{\frac{u}{n}} + 2c(\eta) \frac{\|K\|_2 u}{n\sqrt{\ell}} \right) u^{-1} du \\ &\leq 2 \sum_{\ell \in H} e^{-W_\ell} \int_0^{+\infty} f^2(u) e^{-u} u^{-1} du \\ &\leq C_\eta \sum_{\ell \in H} e^{-W_\ell} \left( \|h\|_\infty \|K\|_1^2 \int_0^{+\infty} e^{-u} du + \frac{\|K\|_2^2}{\ell^2} \int_0^{+\infty} u e^{-u} du \right) \times \frac{1}{n} \\ &\leq C_\eta \sum_{\ell \in H} e^{-W_\ell} \left( \|h\|_\infty \|K\|_1^2 + \frac{\|K\|_2^2}{\ell^2} \right) \times \frac{1}{n}. \end{aligned} \tag{53}$$

We need to choose  $W_\ell$  and  $\eta$  such that

$$\mathbb{E} [\Sigma_n^2] = \mathbb{E} \left[ \sup_{\ell \in H} \left\{ \|Z_\ell\|_2 - \frac{(1+\epsilon)\|K\|_2}{\sqrt{n\ell}} \right\}_+^2 \right] \leq \Lambda. \tag{54}$$



Let  $\theta > 0$ , we choose

$$W_\ell = \frac{\theta^2 \|K\|_2^2}{2\|h\|_\infty \|K\|_1^2 \sqrt{\ell}},$$

the we have

$$\Psi_\ell = (1 + \eta) \frac{\|K\|_2}{\sqrt{n\ell}} + \frac{\theta \|K\|_2}{\sqrt{2n\sqrt{\ell}}} + \frac{c(\eta)\theta^2 \|K\|_2^3}{\|h\|_\infty \|K\|_1^2} \frac{1}{n\ell}.$$

Obviously, the series in (53) is finite and for any  $\ell \in H$ , since  $\ell \leq 1$ , we have

$$\begin{aligned} \Psi_\ell &\leq (1 + \eta + \theta) \frac{\|K\|_2}{\sqrt{n\ell}} + \frac{c(\eta)\theta^2 \|K\|_2^3}{\|h\|_\infty \|K\|_1^2} \frac{1}{n\ell} \\ &\leq \left( 1 + \eta + \theta + \frac{c(\eta)\theta^2 \|K\|_2^2}{\|h\|_\infty \|K\|_1^2} \frac{1}{\sqrt{n\ell}} \right) \frac{\|K\|_2}{\sqrt{n\ell}}. \end{aligned}$$

Since  $H \subset \{\Delta^{-1}, \Delta = 1, \dots, \Delta_{\max}\}$ , if we choose  $\Delta_{\max} = \lfloor \delta n \rfloor$  for some  $\delta > 0$ , then  $\ell_{\min} = \Delta_{\max}^{-1}$  and we obtain

$$\Psi_\ell \leq \left( 1 + \eta + \theta + \frac{c(\eta)\theta^2 \|K\|_2^2 \sqrt{\delta}}{\|h\|_\infty \|K\|_1^2} \right) \frac{\|K\|_2}{\sqrt{n\ell}}.$$

It remains to choose  $\eta = \epsilon/2$  and  $\theta$  small enough such that

$$\theta + \frac{c(\eta)\theta^2 \|K\|_2^2 \sqrt{\delta}}{\|h\|_\infty \|K\|_1^2} = \frac{\epsilon}{2},$$

then

$$\Psi_\ell \leq (1 + \epsilon) \frac{\|K\|_2}{\sqrt{n\ell}},$$

and we get

$$\mathbb{E} [\Sigma_n^2] \leq C_* \times \frac{1}{n},$$

where  $C_*$  is a constant depending on  $\delta, \epsilon, \|h\|_\infty, \|K\|_1$  and  $\|K\|_2$ . Hence, we get (52).

Combining (49) and (52), we obtain

$$\mathbb{E} [\|\hat{h} - h\|_2^2] \leq C_1 \left( \|K_\ell \star h - h\|_2^2 + \frac{\|K\|_2^2}{\ell} \mathbb{E} \left[ \frac{1}{M_T} \right] \right) + C_* \mathbb{E} \left[ \frac{1}{M_T} \right]$$

Moreover, since  $N_T > N_0$ , we have

$$\begin{aligned} \mathbb{E} \left[ \frac{1}{M_T} \right] &= \mathbb{E} \left[ \frac{1}{N_T - N_0} \right] = \mathbb{E} \left[ \frac{N_T}{N_T - N_0} \frac{1}{N_T} \right] = \mathbb{E} \left[ \frac{1}{1 - \frac{N_0}{N_T}} \frac{1}{N_T} \right] \\ &\leq \mathbb{E} \left[ \frac{1}{1 - \frac{N_0}{N_0+1}} \frac{1}{N_T} \right] \\ &\leq (N_0 + 1) \mathbb{E} \left[ \frac{1}{N_T} \right]. \end{aligned} \tag{55}$$

Then, using (10), (12) and (55), recall the definition of  $\varrho(T)^{-1}$  in (29), we obtain for any  $\ell \in H$

$$\mathbb{E} [\|\hat{h} - h\|_2^2] \leq C_1 \left( \|K_\ell \star h - h\|_2^2 + \frac{\|K\|_2^2}{\ell} \varrho(T)^{-1} \right) + C_2 \varrho(T)^{-1}.$$

This ends the proof of Theorem 2.

#### 4.7 Proof of Theorem 3

We begin with the bias term  $\|K_\ell \star h - h\|_2$  in the right hand side of the oracle inequality (30). For any  $\ell \in H$  and  $\gamma \in (0, 1)$ , let  $k = \lfloor \beta \rfloor$  and  $b(\gamma) = K_\ell \star h(\gamma) - h(\gamma)$ , then we have

$$h(\gamma + u\ell) = h(\gamma) + h'(\gamma)u\ell + \cdots + \frac{(u\ell)^k}{(k-1)!} \int_0^1 (1-\theta)^{k-1} h^{(k)}(\gamma + \theta u\ell) d\theta.$$

Since  $K$  is a kernel of order  $\beta^*$  and  $\beta \in (0, \beta^*)$ , we get

$$b(\gamma) = \int K(u) \frac{(u\ell)^k}{(k-1)!} \left[ \int_0^1 (1-\theta)^{k-1} \left( h^{(k)}(\gamma + \theta u\ell) - h^{(k)}(\gamma) \right) d\theta \right] du.$$

Setting  $E_{k,\ell}(u) = |K(u)| \frac{|u\ell|^k}{(k-1)!}$  for the sake of notation. Since  $h \in \mathcal{H}(\beta, L)$  and applying twice the generalized Minskowki's inequality, we obtain

$$\begin{aligned} \|h - \mathbb{E}[\hat{h}]\|_2^2 &= \int b^2(\gamma) d\gamma \\ &\leq \int \left( \int E_{k,\ell}(u) \left[ \int_0^1 (1-\theta)^{k-1} |h^{(k)}(\gamma + \theta u\ell) - h^{(k)}(\gamma)| d\theta \right] du \right)^2 d\gamma \\ &\leq \left( \int E_{k,\ell}(u) \left[ \int \left( \int_0^1 (1-\theta)^{k-1} |h^{(k)}(\gamma + \theta u\ell) - h^{(k)}(\gamma)| d\theta \right)^2 d\gamma \right]^{1/2} du \right)^2 \\ &\leq \left( \int E_{k,\ell}(u) \left[ \int_0^1 (1-\theta)^{k-1} \left( \int |h^{(k)}(\gamma + \theta u\ell) - h^{(k)}(\gamma)|^2 d\gamma \right)^{1/2} d\theta \right] du \right)^2 \\ &\leq \left( \int E_{k,\ell}(u) \left[ \int_0^1 (1-\theta)^{k-1} L(\theta u\ell)^{\beta-k} d\theta \right] du \right)^2 \\ &\leq \left( \int |K(u)| \frac{|u\ell|^k}{(k-1)!} \left[ \int_0^1 (1-\theta)^{k-1} L(u\ell)^{\beta-k} d\theta \right] du \right)^2 \\ &\leq C_{K,L,\beta} \ell^{2\beta}, \end{aligned}$$

where  $C_{K,L,\beta} = \left( \frac{L}{k!} \int |u|^\beta |K(u)| du \right)^2$ .

Finally, we have

$$\mathbb{E} \left[ \|\hat{h} - h\|_2^2 \right] \leq C_1 \inf_{\ell \in H} \left\{ C_{K,L,\beta} \ell^{2\beta} + \frac{\|K\|_2^2}{\ell} \varrho(T)^{-1} \right\} + C_2 \varrho(T)^{-1}. \quad (56)$$

Taking the derivative of the expression inside the  $\inf_{\ell \in H}$  of (56) with respect to  $\ell$ , we obtain the minimizer

$$\ell^* = \left( \frac{\|K\|_2^2}{2\beta C_{K,L,\beta}} \right)^{\frac{1}{2\beta+1}} \varrho(T)^{-\frac{1}{2\beta+1}}.$$

Since the optimal bandwidth  $\hat{\ell}$  is proportional to  $\ell^*$  up to a multiplicative constant. Therefore, by substituting  $\ell$  by  $\hat{\ell}$  in the right hand side of (56), we obtain

$$\mathbb{E} \left[ \|\hat{h} - h\|_2^2 \right] \leq C_3 \varrho(T)^{-\frac{2\beta}{2\beta+1}},$$

with  $C_3$  a constant depending on  $N_0, \delta, \epsilon, \|K\|_1, \|K\|_2, \|h\|_\infty, \beta$  and  $L$ . This ends the proof of Theorem 3.

#### 4.8 Proof of Theorem 4

For  $T > 0$ , let us denote by  $\hat{h}_T$  the estimator of  $h$ . To prove the Theorem 4, we apply the general reduction scheme proposed by Tsybakov [36] (Section 2.2, p.79). We will show the existence of a family  $\mathcal{H}_{m,T} = \{h_{j,T} : j = 0, 1, \dots, m\}$  such that:

- 1)  $h_{j,T} \in \mathcal{H}(\beta, L), j = 0, \dots, m$ .
- 2)  $\|h_{j,T} - h_{k,T}\|_2 \geq 2c e^{-\frac{\beta}{2\beta+1}RT}, 0 \leq j < k \leq m$ .
- 3)  $\frac{1}{m} \sum_{j=1}^m K(P_j, P_0) \leq \vartheta \log(m)$  for  $0 < \vartheta < 1/8$ .  $P_j$  and  $P_0$  are the distribution of observations when the division kernels are  $h_{j,T}$  and  $h_0$ , respectively.  $K(P, Q)$  denotes the Kullback-Leibler divergence between two measures  $P$  and  $Q$ :

$$K(P, Q) = \begin{cases} \int \log \frac{dP}{dQ} dP, & \text{if } P \ll Q \\ +\infty, & \text{otherwise.} \end{cases}$$

Under the preceding conditions 1, 2, 3, Tsybakov [36] (Theorem 2.5, p.99) show that

$$\inf_{\hat{h}_T} \max_{h \in \mathcal{H}_{m,T}} \mathbb{P} \left( \|\hat{h}_T - h\|_2^2 \geq c^2 e^{-\frac{2\beta}{2\beta+1}RT} \right) \geq C', \quad (57)$$

where the infimum is taken over all estimators  $\hat{h}_T$  and positive constant  $C'$  is independent of  $T$ . This will be sufficient to obtain Theorem 4 by [36, Theorem 2.7]. The proof ends with proposing a family  $\mathcal{H}_{m,T}$  and checking the assumptions 1, 2, 3.

##### Construction of the family $\mathcal{H}_{m,T}$ :

The idea is construct a family of perturbations around  $h_0$  which is a symmetric density with respect to  $\frac{1}{2}$  and belongs to  $\mathcal{H}(\frac{L}{2}, \beta)$ . For the simplification, we choose  $h_0(\gamma) = \mathbb{1}_{(0,1)}(\gamma)$ .

Let  $c_0 > 0$  be a real number, and let  $\gamma \in (0, 1)$ ,  $f(\gamma) = LD^{-\beta}g(D\gamma)$  where  $g$  is a regular function having support  $(0, 1)$  and  $\int g(\gamma)d\gamma = 0$ ,  $g \in \mathcal{H}(\frac{1}{2}, \beta)$ , we define

$$D = \lceil c_0 e^{\frac{RT}{2\beta+1}} \rceil \quad \text{and} \quad f_k(\gamma) = f \left( \gamma - \frac{(k-1)}{D} \right),$$

By definition, the functions  $f_k$ 's have disjoint support and one can check that the functions  $f_k \in \mathcal{H}(\frac{L}{2}, \beta)$ .

Then, the function  $h_{j,T}$  will be chosen in

$$\mathcal{D} = \left\{ h_\delta(\gamma) = h_0(\gamma) + c_1 \sum_{k=1}^D \delta_k f_k(\gamma) : \delta = (\delta_1, \dots, \delta_D) \in \{0, 1\}^D \right\},$$

where

$$c_1 = \min \left( \frac{1}{LD^{-\beta}\|g\|_\infty}, 1 \right). \quad (58)$$

We now check that  $h_\delta$  is a density, since  $\int h_\delta(\gamma) d\gamma = \int h_0(\gamma) d\gamma = 1$ , it remains to verify that  $h_\delta(\gamma) \geq 0 \forall \gamma$ . We have

$$\begin{aligned} \inf_{(0,1)} h_\delta(\gamma) &\geq \inf_{(0,1)} h_0 - \|c_1 \sum_{k=1}^D \delta_k f_k\|_\infty \\ &\geq 1 - c_1 L D^{-\beta} \max_k \sup_\gamma |\delta_k| g(D\gamma - (k-1)) \\ &\geq 1 - c_1 L D^{-\beta} \|g\|_\infty \geq 0, \end{aligned}$$

by the choice of  $c_1$ . Thus the family of densities  $\mathcal{D}$  is well-defined.

**1) The condition**  $h_{j,T} \in \mathcal{H}(\beta, L)$ :

Let us denote  $q = \lfloor \beta \rfloor$ , then for all  $\gamma, \gamma' \in (0, 1)$  we have

$$\begin{aligned} |h_\delta^{(q)}(\gamma) - h_\delta^{(q)}(\gamma')| &= |h_0^{(q)}(\gamma) - h_0^{(q)}(\gamma') + c_1 \sum_{k=1}^D \delta_k (f_k^{(q)}(\gamma) - f_k^{(q)}(\gamma'))| \\ &\leq c_1 \sum_{k=1}^D |\delta_k| |f_k^{(q)}(\gamma) - f_k^{(q)}(\gamma')| \\ &\leq c_1 \max_k |f_k^{(q)}(\gamma) - f_k^{(q)}(\gamma')| \\ &\leq c_1 L D^{-\beta} \max_k D^q |g^{(q)}(D\gamma - (k-1)) - g^{(q)}(D\gamma' - (k-1))| \\ &\leq c_1 L D^{\lfloor \beta \rfloor - \beta} D^{\beta - \lfloor \beta \rfloor} |\gamma - \gamma'|^{\beta - \lfloor \beta \rfloor} \leq L |\gamma - \gamma'|^{\beta - \lfloor \beta \rfloor}, \end{aligned}$$

which is always satisfied with  $c_1 = \min\left(\frac{1}{L D^{-\beta} \|g\|_\infty}, 1\right)$ , thus  $h_\delta \in \mathcal{H}(L, \beta)$ .

**2) The condition**  $\|h_{j,T} - h_{k,T}\|_2 \geq 2c e^{-\frac{\beta}{2\beta+1} RT}$ :

For all  $\delta, \delta' \in \{0, 1\}^D$ , we have

$$\begin{aligned} \|h_\delta - h_{\delta'}\|_2 &= \left[ \int_0^1 (h_\delta(\gamma) - h_{\delta'}(\gamma))^2 d\gamma \right]^{1/2} = \left[ \int_0^1 \left( c_1 \sum_{k=1}^D (\delta_k - \delta'_k) f_k(\gamma) \right)^2 d\gamma \right]^{1/2} \\ &= c_1 \left[ \int_0^1 \sum_{k=1}^D (\delta_k - \delta'_k)^2 f_k^2(\gamma) d\gamma \right]^{1/2} = c_1 \left[ \sum_{k=1}^D (\delta_k - \delta'_k)^2 \int_{\frac{k-1}{D}}^{\frac{k}{D}} f_k^2(\gamma) d\gamma \right]^{1/2} \\ &= c_1 \left[ \sum_{k=1}^D (\delta_k - \delta'_k)^2 \int_{\frac{k-1}{D}}^{\frac{k}{D}} L^2 D^{-2\beta} g^2(D\gamma - (k-1)) d\gamma \right]^{1/2} \\ &= c_1 L D^{-\beta-1/2} \|g\|_2 \left[ \sum_{k=1}^D (\delta_k - \delta'_k)^2 \right]^{1/2} = c_1 L D^{-\beta-1/2} \|g\|_2 \sqrt{d_H(\delta, \delta')}, \end{aligned}$$

where  $d_H(\delta, \delta') = \sum_{k=1}^D \mathbb{1}\{\delta_k \neq \delta'_k\}$  is the Hamming distance between  $\delta$  and  $\delta'$ .

According to the Lemma of Varshamov-Gilbert (cf. Tsybakov [36], p.104), there exist a subset  $\{\delta^{(0)}, \dots, \delta^{(m)}\}$  of  $\{0, 1\}^D$  with cardinal (59) such that  $\delta^{(0)} = (0, \dots, 0)$ ,

$$m \geq 2^{D/8}, \quad (59)$$

and

$$d_H(\delta^{(j)}, \delta^{(k)}) \geq \frac{D}{8}, \forall 0 \leq j < k \leq m. \quad (60)$$

Then, by setting  $h_{j,T}(x) = h_{\delta^{(j)}}(x)$ ,  $j = 0, \dots, m$ , we obtain

$$\begin{aligned} \|h_{j,T} - h_{k,T}\|_2 &= c_1 L D^{-\beta-\frac{1}{2}} \|g\|_2 \sqrt{d_H(\delta^{(j)}, \delta^{(k)})} \\ &\geq c_1 L D^{-\beta-1/2} \|g\|_2 \sqrt{\frac{D}{8}} \\ &\geq \frac{c_1 L}{4} \|g\|_2 D^{-\beta} \end{aligned}$$

whenever  $D \geq 8$ .

Suppose that  $N_T \geq N_{T^*}$  where  $T^* = \log\left(\frac{7}{c_0}\right) \frac{2\beta+1}{R}$ . Then,  $D \geq 8$  and  $D^\beta \leq (2c_0)^\beta e^{\frac{\beta}{2\beta+1}RT}$ . This implies:

$$\|h_{j,T} - h_{k,T}\|_2 \geq \frac{c_1 L}{4} \|g\|_2 (2c_0)^{-\beta} e^{-\frac{\beta}{2\beta+1}RT},$$

But,

$$\min\left(\frac{1}{L\|g\|_\infty}, 1\right) \leq c_1 \leq 1$$

Hence, we obtain

$$\|h_{j,T} - h_{k,T}\|_2 \geq 2c e^{-\frac{\beta}{2\beta+1}RT},$$

where

$$c = \frac{\min(1, L\|g\|_2)}{8} (2c_0)^{-\beta}.$$

**3) The condition  $\frac{1}{m} \sum_{j=1}^m K(P_j, P_0) \leq \vartheta \log(m)$  for  $0 < \vartheta < 1/8$ :**

We need to show that for all  $\delta \in \{0, 1\}^D$ ,

$$K(P_\delta, P_0) \leq \vartheta \log(m),$$

where

$$K(P_\delta, P_0) = \mathbb{E} \left[ \log \frac{dP_\delta}{dP_0} |_{\mathcal{F}_T}(Z) \right],$$

and where  $(Z_t)_{t \in [0, T]}$  is defined in (5) with the random measure  $Q$  having intensity  $q(ds, di, d\gamma) = Rh_\delta(\gamma) ds n(di) d\gamma$ .

Here, the difficulty comes from the fact that  $N_T$  is variable because the observations result from a stochastic process  $Z_t$ . The law of these observations is not a probability distribution on a fixed  $\mathbb{R}^n$  where  $n$  would be the sample size, but rather a probability distribution on a path space.  $P_\delta$  is the probability distribution when the Poisson point measure  $Q$  has intensity  $Rh_\delta(\gamma) ds n(di) d\gamma$ . Thus a natural tool is to use Girsanov's theorem (see [21], Theorem 3.24, p. 159) saying that  $P_\delta$  is absolutely continuous with respect to  $P_0$  on  $\mathcal{F}_T$  with

$$\frac{dP_\delta}{dP_0} |_{\mathcal{F}_T} = \mathfrak{D}_T^\delta,$$

where  $(\mathfrak{D}_t^\delta)_{t \in [0, T]}$  is the unique solution of the following SDE (see Proposition 4.17 of [34] for a similar SDE):

$$\mathfrak{D}_T^\delta = 1 + \int_0^T \int_{\mathcal{E}} \mathfrak{D}_{s-}^\delta \mathbf{1}_{\{i \leq N_{s-}\}} \left( \frac{h_\delta(\gamma)}{h_0(\gamma)} - 1 \right) Q(ds, di, d\gamma). \quad (61)$$

Apply Itô formula for jump processes to (61), we get

$$\begin{aligned}\log \mathfrak{D}_T^\delta &= \int_0^T \int_{\mathcal{E}} \mathbf{1}_{\{i \leq N_{s-}\}} \left[ \log \left( \mathfrak{D}_{s-}^\delta - \left( \frac{h_\delta(\gamma)}{h_0(\gamma)} - 1 \right) \mathfrak{D}_{s-}^\delta \right) - \log \mathfrak{D}_{s-}^\delta \right] Q(ds, di, d\gamma) \\ &= \int_0^T \int_{\mathcal{E}} \mathbf{1}_{\{i \leq N_{s-}\}} \log \frac{h_\delta(\gamma)}{h_0(\gamma)} Q(ds, di, d\gamma) = \sum_{i=1}^{N_T} \log \frac{h_\delta(\Gamma_i^1)}{h_0(\Gamma_i^1)}\end{aligned}$$

by definition of  $(\Gamma_1^1, \dots, \Gamma_{N_T}^1)$ .

Then,

$$\begin{aligned}K(P_\delta, P_0) &= \mathbb{E}_\delta [\log \mathfrak{D}_T^\delta] = \mathbb{E}_\delta \left[ \sum_{i=1}^{N_T} \log \frac{h_\delta(\Gamma_i^1)}{h_0(\Gamma_i^1)} \right] \\ &= \mathbb{E}[N_T] \mathbb{E}_\delta \left[ \log \frac{h_\delta(\Gamma_1^1)}{h_0(\Gamma_1^1)} \right] = \mathbb{E}[N_T] \int_0^1 h_\delta(\gamma) \log \frac{h_\delta(\gamma)}{h_0(\gamma)} d\gamma.\end{aligned}$$

Here,  $\mathbb{E}[N_T]$  does not depend on  $h_\delta$  and we have  $\mathbb{E}[N_T] = N_0 e^{RT}$ . Thus, recall the definition of  $h_\delta(\cdot)$  and note that  $\log(1+x) \leq x$  for  $x > -1$ , we get

$$\begin{aligned}K(P_\delta, P_0) &= N_0 e^{RT} \int_0^1 h_\delta(\gamma) \log(h_\delta(\gamma)) d\gamma \\ &= N_0 e^{RT} \int_0^1 \left( 1 + c_1 \sum_{k=1}^D \delta_k f_k(\gamma) \right) \log \left( 1 + c_1 \sum_{k=1}^D \delta_k f_k(\gamma) \right) d\gamma \\ &= N_0 e^{RT} \sum_{k=1}^D \int_{\frac{k-1}{D}}^{\frac{k}{D}} (1 + c_1 \delta_k f_k(\gamma)) \log(1 + c_1 \delta_k f_k(\gamma)) d\gamma \\ &= N_0 e^{RT} \sum_{k=1}^D \delta_k \int_0^{1/D} (1 + c_1 f(\gamma)) \log(1 + c_1 f(\gamma)) d\gamma \\ &\leq N_0 e^{RT} D \int_0^{1/D} (1 + c_1 f(\gamma)) c_1 f(\gamma) d\gamma \\ &\leq N_0 e^{RT} \left[ c_1 L D^{-\beta} \int_0^{1/D} g(D\gamma) D d\gamma + c_1^2 L^2 D^{-2\beta} \int_0^{1/D} g^2(D\gamma) D d\gamma \right] \\ &\leq N_0 e^{RT} c_1^2 L^2 D^{-2\beta} \int_0^1 g^2(\gamma) d\gamma \\ &\leq N_0 c_1^2 L^2 \|g\|_2^2 e^{RT} c_0^{-2\beta} e^{-\frac{2\beta}{2\beta+1} RT} \\ &\leq N_0 L^2 \|g\|_2^2 c_0^{-2\beta-1} D \quad \text{since } c_1 \leq 1.\end{aligned}$$

From (59), we have  $m \geq 2^{D/8}$  then

$$D \leq \frac{8 \log(m)}{\log(2)}.$$

Hence, if we set

$$c_0 = \left( \frac{8 N_0 L^2 \|g\|_2^2}{\vartheta \log(2)} \right)^{1/(2\beta+1)},$$

we obtain  $K(P_\delta, P_0) \leq \vartheta \log(m)$ . This ends the proof of Theorem 4.

## Acknowledgements

The author is deeply grateful to V.C. Tran, V. Rivoirard and T.M. Pham Ngoc for the guidance and useful suggestions. The author would like to express his sincere thanks to P. Massart for constructive comments that improved the results. The author would like to thank Centre de Ressources Informatiques de Lille 1 (CRI) for the computational cluster to implement the numerical simulations. This work is supported by Program 911 of Vietnam Ministry of Education and Training and is partly supported by the French Agence Nationale de la Recherche (ANR 2011 BS01 010 01 projet Calibration).

## References

- [1] H. Aguilaniu, L. Gustafsson, M. Rigoulet, T. Nyström. Asymmetric Inheritance of Oxidatively Damaged Proteins During Cytokinesis. *Science*, 299, 1751, 2003.
- [2] M. Ackermann, S. C. Stearns, U. Jenal. Senescence in a Bacterium with Asymmetric Division. *Science*, 300, 1920, 2003.
- [3] K.B. Athreya and P.E. Ney. *Branching Processes*. Springer edition, 1970.
- [4] V. Bansaye. Proliferating parasites in dividing cells: Kimmels branching model revisited. *The Annals of Applied Probability*, (2008): 967-996.
- [5] V. Bansaye, J.-F. Delmas, L. Marsalle, and V. C. Tran. Limit theorems for markov processes indexed by continuous time galton-watson trees. *The Annals of Applied Probability*, 21, 2011.
- [6] V. Bansaye and V. C. Tran. Branching feller diffusion for cell division with parasite infection. *ALEA, Lat. Am. J. Probab. Math. Stat.*, 8, 2011.
- [7] V. Bansaye, J. C. Pardo and C. Smadi. On the extinction of continuous state branching processes with catastrophes. *Electron. J. Probab.*, 18.106 (2013): 1-31.
- [8] B. Bercu, B. De Saporta and A. Gégout-Petit. Asymptotic analysis for bifurcating autoregressive processes via a martingale approach. *Electron. J. Probab.*, 14.87 (2009): 2492-2526.
- [9] K. Bertin, C. Lacour, and V. Rivoirard. Adaptive pointwise estimation of conditional density function. *to appear in the Ann. Inst. H. Poincaré Probab. Statist.*, 2015.
- [10] S. Valère Bitseki Penda. Deviation inequalities for bifurcating Markov chains on GaltonWatson tree. *ESAIM: Probability and Statistics*, 19 (2015): 689-724.
- [11] B. Cloez. Limit theorems for some branching measure-valued processes. *hal-00598030*, 2011.
- [12] J.-F. Delmas and L. Marsalle. Detection of cellular aging in a Galton-Watson process. *Stochastic Processes and their Application*, 120 (12), 2495-2519 (2010).
- [13] M. Doumic, M. Hoffmann, N. Krell, and L. Robert. Statistical estimation of a growth-fragmentation model observed on a genealogical tree. *arXiv:1210.3240*, 2012.
- [14] M. Doumic, M. Hoffmann, P. Reynaud-Bouret, and V. Rivoirard. Nonparametric estimation of the division rate of a size-structured population. *SIAM Journal on Numerical Analysis*, 50, 2012.
- [15] S. N. Evans and D. Steinsaltz. Damage segregation at fissioning may increase growth rates: A superprocess model. *Theoretical Population Biology*, 71, 473-490, 2007.
- [16] A. Goldenshluger and O. Lepski. Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *The Annals of Statistics*, 39(3):1608-1632, 2011.
- [17] J. Guyon. Limit theorems for bifurcating Markov chains. Application to the detection of cellular aging. *The Annals of Applied Probability*, 17.5/6 (2007): 1538-1569.
- [18] T.E. Harris. *The Theory of Branching Processes*. Springer, Berlin, 1963.

- [19] M. Hoffmann and A. Olivier. Nonparametric estimation of the division rate of an age dependent branching process. *arXiv:1412.5936*, 2014.
- [20] N. Ikeda and S. Watanabe. *Stochastic differential equations and diffusion processes*. volume 24 of North-Holland Mathematical Library.
- [21] J. Jacob and A. N. Shiryaev. *Limit Theorems for Stochastic Processes*. Springer-Verlag, Berlin, 1987.
- [22] C-Y. Lai, E. Jaruga, C. Borghouts and S. M. Jazwinski. A Mutation in the ATP2 Gene Abrogates the Age Asymmetry Between Mother and Daughter Cells of the Yeast *Saccharomyces cerevisiae*. *Genetics*, 162, 73-87, 2002.
- [23] C. Lacour and P. Massart. Minimal penalty for goldenshluger-lepski method. *arXiv:1503.00946v1*, (4:3), 2015.
- [24] A.B. Lindner, R. Madden, A. Demarez, E.J. Stewart, F. Taddei, Asymmetric segregation of protein aggregates is associated with cellular aging and rejuvenation. *PNAS*, (105:8), 3076-3081, 2015.
- [25] P. Massart. *Concentration Inequalities and Model Selection*. Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, 6-23, 2003. Springer, 2007.
- [26] S. P. Meyn and R. L. Tweedie. Stability of markovian processes iii: Foster - lyapunov criteria for continuous-time processes. *Advances in Applied Probability*, 25, 1993.
- [27] J. B. Moseley. Cellular Aging: Symmetry Evades Senescence. *Current Biology*, Volume 23, Issue 19, R871 – R873, 2013.
- [28] J. Peter. A general stochastic model for population development. *Scandinavian Actuarial Journal*, 1969(1-2):84–103, 1969.
- [29] Division in *Escherichia coli* is triggered by a size-sensing rather than a timing mechanism. *BMC Biology*, 12:17, 2014.
- [30] Sheldon M. Ross. *Stochastic Processes*. Wiley, second edition, 1995.
- [31] P. Reynaud-Bouret, V. Rivoirard V, F. Grammont, and C. Tuleau-Malot. Goodness-of-fit tests and nonparametric adaptive estimation for spike train analysis. *Journal of Mathematical Neuroscience*, (4:3), 2014.
- [32] B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, London New York, 1986, 175 pp.
- [33] E. J. Stewart, R. Madden, G. Paul, and F. Taddei. Aging and Death in an Organism That Reproduces by Morphologically Symmetric Division. *PLOS Biology*, 3, 2005.
- [34] V. C. Tran. *Modèles particuliers stochastiques pour des problèmes d'évolution adaptative et pour l'approximation de solutions statistiques*. PhD dissertation, Université Paris X - Nanterre, <http://tel.archives-ouvertes.fr/tel-00125100>, 2006.
- [35] V. C. Tran. Large population limit and time behaviour of a stochastic particle model describing an age-structured population. *ESAIM: Probability and Statistics*, 12, 2008.
- [36] A. B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer series in Statistics. Springer, 2004.